

AD \_\_\_\_\_

Award Number DAMD17-94-J-4063

TITLE: c-MYC Protein-Protein and Protein-DNA Interactions:  
Targets for Therapeutic Intervention

PRINCIPAL INVESTIGATOR: T. Keith Blackwell, M.D., Ph.D.

CONTRACTING ORGANIZATION: Center for Blood Research  
Boston, Massachusetts 02115

REPORT DATE: September 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20000907 130

DTIC QUALITY INSPECTED 4

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

<b>1. AGENCY USE ONLY (Leave blank)</b>		<b>2. REPORT DATE</b> September 1999	<b>3. REPORT TYPE AND DATES COVERED</b> Final (1 Sep 94 - 31 Aug 99)	
<b>4. TITLE AND SUBTITLE</b> C-MYC Protein-Protein and Protein-DNA Interactions: Targets for Therapeutic Intervention			<b>5. FUNDING NUMBERS</b> DAMD17-94-J-4063	
<b>6. AUTHOR(S)</b> T. Keith Blackwell, M.D., Ph.D.				
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Center for Blood Research Boston, Massachusetts 02115 E-MAIL:			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			<b>10. SPONSORING / MONITORING AGENCY REPORT NUMBER</b>	
<b>11. SUPPLEMENTARY NOTES</b>				
<b>12a. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release; distribution unlimited				<b>12b. DISTRIBUTION CODE</b>
<b>13. ABSTRACT (Maximum 200 Words)</b>  <p>Basic-helix-loop-helix (bHLH) transcription factor proteins are involved in cell growth and differentiation, and the bHLH protein c-Myc has been implicated in breast cancer. We have determined that the sequence preference and specificity with which particular bHLH protein subclasses bind DNA are determined by binding conformations, which in turn are dictated by certain bHLH DNA binding region amino acids that are critical for protein function, but not for binding affinity. The data identify subtle effects on DNA binding that have important biological consequences, and reveal mechanisms by which different protein sub-family members can bind to different sites that share the CANNTG consensus.</p> <p>The zinc finger protein TTP is induced by many growth factors, and is required for regulation of TNF<math>\alpha</math> expression. We have shown that continuous expression of TTP or the related proteins TIS11b and TIS11d induces apoptosis and affects growth regulatory pathways. Like c-Myc, TTP also sensitizes cells to apoptosis induced by TNF<math>\alpha</math>. Our data suggest that TTP is important in growth factor signaling, and that its role in TNF<math>\alpha</math> function is likely to be complex.</p>				
<b>14. SUBJECT TERMS</b> c-Myc                      oncogenes      apoptosis helix-loop-helix      TNF                      growth factor			<b>15. NUMBER OF PAGES</b> 85	
			<b>16. PRICE CODE</b>	
<b>17. SECURITY CLASSIFICATION OF REPORT</b> Unclassified	<b>18. SECURITY CLASSIFICATION OF THIS PAGE</b> Unclassified	<b>19. SECURITY CLASSIFICATION OF ABSTRACT</b> Unclassified	<b>20. LIMITATION OF ABSTRACT</b> Unlimited	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

☒ Where copyrighted material is quoted, permission has been obtained to use such material.

☒ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

☒ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

X In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

T. G. H. H.

PI - Signature

12/8/99

Date

**TABLE OF CONTENTS:**

<u>Report Documentation Page:</u>	p 1
<u>Foreword:</u>	p 2
<u>Table of Contents:</u>	p 3
<u>Introduction:</u>	pp 4-8
<u>Body:</u>	pp 8-15
<u>Key Research Accomplishments:</u>	p 15
<u>Reportable Outcomes:</u>	pp 15-16
<u>Conclusions:</u>	pp 16-24
<u>References:</u>	pp 24-31
<u>Appendix (Figures):</u>	pp 32-51
<u>Final Report Bibliography and Personnel:</u>	p 52

## INTRODUCTION:

### A. Foreword:

As described in the previous report, and in accordance with the approved amended statement of work for this project, we have concentrated our efforts on understanding basic-helix-loop-helix protein DNA binding specificity, and on elucidating the functions of TIS11/TTP proteins. Each section below will be divided into two parts, each of which will correspond to one of these projects.

### B. Determinants of bHLH protein DNA binding specificity:

A large family of transcriptional regulators is defined by the basic-helix-loop-helix (bHLH) motif (Murre et al., 1989), in which a DNA-binding basic region (BR) lies immediately amino terminal to the HLH dimerization segment (Davis et al., 1990; Murre et al., 1989; Voronova and Baltimore, 1990). In metazoans, bHLH proteins are involved in specification of multiple cell types (Lee, 1997; Olson and Klein, 1994; Weintraub et al., 1991). Some bHLH family members function as homodimers, but others appear to act together with a heterodimeric partner (Weintraub et al., 1991). For example, the closely-related bHLH proteins that mediate myogenic differentiation, including MyoD, are thought to function as heterodimers with E proteins, a widely-expressed bHLH protein subgroup that is exemplified by the E2A proteins (Chakraborty et al., 1991; Davis et al., 1990; Lassar et al., 1991; Neuhold and Wold, 1993). Most bHLH protein dimers bind to the consensus CANNTG (the E box), with each respective BR binding to a half site (Blackwell and Weintraub, 1990; Ellenberger et al., 1994; Ferre-D' Amare et al., 1993; Ferre-D'Amare et al., 1994; Ma et al., 1994; Parraga et al., 1998; Shimizu et al., 1997). Given the many regulatory processes in which bHLH proteins are involved, the apparent simplicity of the CANNTG consensus raises the important question of how different bHLH proteins act only on appropriate target genes (Weintraub et al., 1991).

In part, the specificity with which bHLH proteins function derives from preferential recognition of different classes of CANNTG sites by different bHLH protein subgroups. The HLH segment consists of a parallel, left-handed, four helix bundle (Fig. 1) (Ellenberger et al., 1994; Ferre-D' Amare et al., 1993; Ferre-D'Amare et al., 1994; Ma et al., 1994; Parraga et al., 1998; Shimizu et al., 1997). The BR is unstructured in solution (Anthony-Cahill et al., 1992), but when bound to DNA it extends N-terminally from the HLH segment as an  $\alpha$ -helix that crosses the major groove (Fig. 1). Crystallographic analyses have revealed some differences in how these proteins bind DNA. For example, in Myc-family and related bHLH proteins, an arginine (Arg) residue at BR position 13 (Fig. 2) specifies recognition of CACGTG sites (Blackwell et al., 1993; Dang et al., 1992; Halazonetis and Kandil, 1992; Van Antwerp et al., 1992) by contacting bases in the center (Ferre-D' Amare et al., 1993; Ferre-D'Amare et al., 1994; Shimizu et al., 1997). However, it still is not understood how bHLH proteins which have a different amino acid at BR position 13 (Fig. 2) bind preferentially to distinct CANNTG sites (Blackwell and Weintraub, 1990; Dang et al., 1992), or how bHLH proteins establish differences in flanking sequence selectivity (Blackwell and Weintraub, 1990; Fisher et al., 1993; Gould and Bresnick, 1998) that can be of biological importance (Aksan and Goding, 1998; Jennings et al., 1999).

Many bHLH proteins that lack R<sub>13</sub>, including MyoD and other E2A partners (Fig. 2), can bind to similar DNA sequences *in vitro* but act on different tissue-specific genes (Weintraub et al., 1991). Cooperative or inhibitory relationships with other

transcriptional regulators might contribute to this specificity (Lemerrier et al., 1998; Molkentin and Olson, 1996; Postigo and Dean, 1999; Weintraub et al., 1994), but it is not likely to derive entirely from other lineage-specific factors, because MyoD can induce myogenesis in many different cell types (Weintraub et al., 1991). Initiation of myogenesis by MyoD and other myogenic bHLH proteins depends upon three residues that are located within the BR and the BR-HLH junction ( $A_5$ ,  $T_6$ , and  $K_{15}$ ; Figs. 1 and 2). These "myogenic" residues are not essential for binding a muscle-specific site *in vitro* or *in vivo*, suggesting that they are involved in other critical interactions (Brennan et al., 1991; Davis et al., 1990; Davis and Weintraub, 1992; Schwarz et al., 1992; Weintraub et al., 1991). These interactions have been proposed to involve distinct co-factors (Brennan et al., 1991; Davis et al., 1990; Weintraub et al., 1991), and the unmasking of an activation domain in MyoD or the myogenic co-factor MEF2 (Bengal et al., 1994; Black et al., 1998; Huang et al., 1998; Weintraub et al., 1991). In the MyoD-DNA structure,  $K_{15}$  is oriented away from the DNA, but  $A_5$  and  $T_6$  face the major groove and could not contact other proteins directly (Ma et al., 1994) (Fig. 1). However, the latter two residues could influence protein-protein interactions indirectly, by affecting how the BR helix is positioned on the DNA (Ma et al., 1994). Although substitutions at these positions might not substantially impair binding to particular CANNTG sites, it is important to determine whether they might have more subtle influences on sequence specificity that could reflect conformational effects.

We have determined that the myogenic residues  $A_5$  and  $T_6$  establish the characteristic MyoD sequence preference, which includes a CAGCTG core. Individual substitutions at these BR positions simultaneously alter preferences for multiple bases that MyoD does not contact directly (Ma et al., 1994), indicating that these preferences are determined indirectly, by how the BR helix is positioned on the DNA. This mechanism is distinct from the standard model for sequence specificity, in which preferred bases are contacted directly (Pabo and Sauer, 1992; Steitz, 1990). The corresponding BR residues are also required for the sequence preferences of E2A proteins, which can recognize either of two distinct half sites depending upon their dimerization partner. E2A homodimers and E2A + MyoD heterodimers bind to asymmetric sites that include a CACCTG core. In contrast, as a heterodimer with the bHLH protein Twist, E2A binds preferentially to half of the symmetric sequence CATATG. The preference of E2A for the former asymmetric sites depends not only upon the BR sequence, but also upon BR positioning that involves the junction region. An analysis of DNA binding by MyoD and E2A junction and BR mutants indicates that a MyoD-like sequence specificity is associated with, but not sufficient for, myogenesis. This supports the model that the BR-junction region is also involved in other critical interactions. The results suggest that E2A and its partner bHLH proteins bind DNA by adopting a limited number of preferred BR conformations, each of which is associated with a characteristic DNA sequence preference. They also predict that binding of co-factors to the MyoD BR might be influenced by how it is positioned on the DNA, and are consistent with the idea that relatively subtle differences in binding sequence recognition can modulate bHLH protein activity *in vivo*.

### **C. TTP/TIS11 proteins and cell survival:**

Complex programs of gene expression ensue when quiescent cells are stimulated by growth factors to enter the cell cycle (Iyer et al., 1999). The "immediate-early" genes are induced directly by many stimuli, and during some apoptotic events. Their products include transcription factors associated with proliferation and apoptosis

(Hafezi et al., 1997; Shi et al., 1992; Zhan et al., 1997). In addition, during these responses the localization, stability, and translation of specific mRNAs are affected (Brown and Schreiber, 1996; Chen and Shyu, 1995), indicating that post-transcriptional regulatory mechanisms are also involved.

The immediate early protein tristetraprolin (TTP; also known as TIS11, Nup475, and G0S24) is induced transiently in various cell types by diverse stimuli, and during regeneration of certain tissues (DuBois et al., 1990; Lai et al., 1990; Varnum et al., 1989; Worthington et al., 1996). TTP is closely related to the TIS11b and TIS11d proteins, particularly within tandem Cys-X<sub>8</sub>-Cys-X<sub>5</sub>-Cys-X<sub>3</sub>-His (Cys<sub>3</sub>His) zinc fingers (Varnum et al., 1991). Each of these three proteins, which we refer to as the TTP/TIS11 proteins, is induced rapidly by multiple different agents, although they vary with respect to their baseline mRNA levels and induction by particular stimuli (Corps and Brown, 1995; Gomperts et al., 1992; Varnum et al., 1991). Other Cys<sub>3</sub>His zinc finger proteins are involved in mRNA binding, cleavage, or processing, or have been implicated in post-transcriptional gene regulation (Barabino et al., 1997; Batchelder et al., 1999; Guedes and Priess, 1997; Murray et al., 1997; Rudner et al., 1998; Tabara et al., 1999; Tronchere et al., 1997) predicting that TTP/TIS11 proteins also perform mRNA-associated functions.

Although TTP is induced in a variety of contexts, mice in which the TTP gene has been disrupted (TTP <sup>-/-</sup> mice) have a limited phenotype. They are normal at birth, but later develop a systemic inflammatory, arthritic, and myeloproliferative syndrome which is mediated by the cytokine tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), and derives from an abnormality in non-lymphoid hematopoietic cells (Carballo et al., 1997; Carballo et al., 1998; Taylor et al., 1996). When stimulated *in vitro*, TTP <sup>-/-</sup> macrophages produce moderately elevated levels of TNF- $\alpha$  protein and mRNA, the half-life of which is prolonged. TNF- $\alpha$  and many cytokine and growth factor-induced mRNAs are regulated post-transcriptionally through AU-rich elements (AREs) in their 3' untranslated regions (Chen and Shyu, 1995; Ross, 1995; Shaw and Kamen, 1986). When TTP is over-expressed, TNF- $\alpha$  and other cytokine mRNAs are deadenylated and destabilized, and binding of TTP to the TNF- $\alpha$  ARE can be detected readily (Carballo et al., 1998; Lai et al., 1999). These observations, and the finding that TTP is induced by TNF- $\alpha$ , have suggested the model that TTP normally destabilizes the TNF- $\alpha$  mRNA directly, through a feedback mechanism (Carballo et al., 1998; Lai et al., 1999). It is intriguing, however, that TNF- $\alpha$  mRNA levels are also decreased by low-level TTP expression but increased by intermediate TTP amounts (Lai et al., 1999), suggesting that TTP may have complex and apparently indirect effects on TNF- $\alpha$  expression.

Although TTP/TIS11 proteins are evolutionarily conserved and induced by numerous extracellular stimuli, suggesting a broader role, no other functions of metazoan TTP/TIS11 proteins have been described. TTP is induced during apoptosis, however, in response to the breast cancer susceptibility protein BRCA1 (Harkin et al., 1999), and withdrawal of growth factors from neuronal cells (Mesner et al., 1995). In *S. pombe*, a related protein is required for effective transmission of a pheromone-induced *ras*/mitogen-activated protein kinase signal (Kano et al., 1995). In addition, a *nim1 cdc25* mutant can be complemented by either the *cdc2* kinase or a TTP/TIS11 gene, suggesting a cell cycle effect (Warbrick and Glover, 1994). A TTP/TIS11-related protein in *S. cerevisiae* is required for normal metabolism, and retards cell growth when overexpressed (Ma and Herschman, 1995; Thompson et al., 1996). These observations indicate that TTP/TIS11 proteins might influence pathways regulatory that regulate survival, differentiation, or proliferation.

If TTP acts on such pathways when it is induced transiently, its continuous expression might be predicted to affect cell growth or viability. Supporting this idea, we report that continuous TTP expression causes various cell types to undergo apoptotic cell death. This response occurs at TTP expression levels which are comparable to those attained transiently during serum stimulation. Each TTP/TIS11 protein stimulates apoptosis with similar frequency and timing, and by various criteria this cell death appears analogous to apoptosis induced by oncoproteins such as E2F-1, or the immediate early protein c-Myc. In addition, TTP differs from TIS11b and TIS11d in that, like both E2F-1 and c-Myc, it appears to sensitize cells to induction of apoptosis by TNF- $\alpha$ . The data indicate that TTP/TIS11 proteins generally act similarly on growth or survival pathways, but also that TTP may have a distinct influence on responses to TNF- $\alpha$ . They suggest that the role of TTP in TNF- $\alpha$  regulation might be complex, and imply that in TTP  $-/-$  mice some TTP functions may be compensated for by other mechanisms.

## **BODY:**

### **A. Determinants of bHLH protein DNA binding specificity:**

**Myogenic BR residues and MyoD DNA binding preferences.** Identification of the myogenic BR residues stemmed originally from studies in which the MyoD BR was replaced with that of E12, a product of the alternatively-spliced E2A gene (Murre et al., 1989). This MyoD mutant (MD(E12B); Fig. 2) binds to a muscle-specific regulatory site as a heterodimer with E2A proteins either *in vitro* or *in vivo*, but cannot induce myogenesis in a cell culture assay or activate transcription through a muscle-specific enhancer (Davis et al., 1990; Weintraub et al., 1991). Re-substitution of the "myogenic" residues A<sub>5</sub> and T<sub>6</sub> (Fig. 2) into MD(E12B) restores its activity in these functional assays (Weintraub et al., 1991). Similar results are obtained when A<sub>5</sub> and T<sub>6</sub> are mutated within MyoD (Davis and Weintraub, 1992; Huang et al., 1998; Weintraub et al., 1991), and when analogous substitutions are made in the context of the myogenic bHLH protein Myogenin (Brennan et al., 1991). These experiments implicate A<sub>5</sub> and T<sub>6</sub> in mechanisms that are of functional importance, but not essential for binding to a particular muscle-specific DNA sequence.

We employed an *in vitro* selection strategy (Blackwell and Weintraub, 1990) to test whether such mutations might have more subtle effects on how MyoD binds specifically to DNA. To identify sequences to which these mutants bind preferentially, we used sequence libraries in which only positions within and flanking the CANNTG consensus are randomized (Fig. 3A), so that the position of bHLH protein binding along the DNA is fixed. This strategy makes it possible to sequence the selected sites as a pool, and thereby to analyze a very large population of selected sites simultaneously (Blackwell et al., 1990; Blackwell and Weintraub, 1990). It reveals the relative preferences for individual bases at each site position, and can detect subtle differences that might not be identified through more conventional approaches.

This assay has previously shown that the preferred MyoD binding consensus is G/AACAGCTGTT/C (Figs. 3B and 3C), and that the E2A proteins E12 and E47 overlap considerably with MyoD in their binding properties, but prefer sites that have an asymmetric CACCTG core sequence (Fig. 3C) (Blackwell and Weintraub, 1990).



However, in contrast to either of these proteins, the MD(E12B) mutant prefers the sequence G/ACCATATGGT/C, which differs from the MyoD preferred site over the 8 central base pairs, and contains the distinct core sequence CATATG (Figs. 3B and 3C). This sequence and related elements are normally targeted by the bHLH protein Twist, an E protein partner that is involved in mesodermal cell fate specification (Cripps et al., 1998; Harfe et al., 1998; Michelson, 1996; Szymanski and Levine, 1995; Yin et al., 1997) (Fig. 2). Back-substitution of A<sub>5</sub> of MyoD into MD(E12B), which is not sufficient for myogenic activity in cell culture assays (Weintraub et al., 1991), results in preferences that are slightly more similar to those of MyoD at positions  $\pm 4$ , (MD(E12B-A); Figs. 2, 3B, and 3C). However, introduction of both A<sub>5</sub> and T<sub>6</sub>, which restores myogenesis (Brennan et al., 1991; Weintraub et al., 1991), results in preferences across the entire site that are indistinguishable from those of MyoD (MD(E12B-AT); Figs. 2, 3B, and 3C).

To determine whether these sequence preferences reflect significant differences in binding affinity and specificity, we compared binding of these proteins to individual oligonucleotides that correspond to the MyoD and Twist preferences, and differ only at positions within and adjacent to the CANNTG consensus (Fig. 3D). Supporting the *in vitro* selection findings, both MyoD and MyoD(E12B-AT) homodimers bound with higher affinity to the preferred MyoD site than to the Twist site (Fig. 3D, lanes 1, 4, 5, and 8). In contrast, the Twist site was preferred by MD(E12B) and, to a lesser extent, MyoD(E12B-A) (Fig. 3D, lanes 2, 3, 6, and 7). In a binding competition assay, specific DNA binding by MD(E12B-AT) was competed much more effectively by the MyoD site (Fig. 4A, lanes 4, 7, 10, 13, and 16), and binding by either MD(E12B) or MD(E12B-A) was competed better by the Twist site (Fig. 4B, lanes 2, 3, 8, 9, 14, and 15). A c-Myc preferred site (CACGTG; not shown) was a relatively poor competitor of binding by each of these proteins (Figs. 4A and B, lanes 17-19). The data show that introduction of A<sub>5</sub> and T<sub>6</sub> into MD(E12B) restores not only myogenic activity (Fig. 2) but also the MyoD DNA binding preference. This substitution affects sequence recognition across 4 bp within each half-site (Figs. 3A and B), indicating a global effect on how the BR helix is positioned on the DNA. The finding that MD(E12B) is distinct from either MyoD or E12 in its binding sequence preference also indicates that DNA recognition by an E2A BR can be profoundly influenced by its molecular context.

**Influence of BR positioning on MyoD/E2A and Twist/E2A heterodimer sequence preferences.** Twist and E2A proteins appear to cooperate *in vivo* to regulate transcription through CATATG sites (Harfe et al., 1998), suggesting that the DNA sequence recognition properties of E2A might be altered by heterodimerization with Twist. However, an alternative possibility is that functional Twist + E2A recognition sites are distinct from their *in vitro* binding preference (Huang et al., 1996). To address this question, we performed *in vitro* selection on Twist + E12 complexes. Twist homodimers and Twist + E12 heterodimers both preferred sites that contain the core sequence CATATG (Figs. 5A and B). They were similar to MD(E12B) and especially to MD(E12B-A) in their preferences at  $\pm 4$ , but selected MyoD-like sequences at  $\pm 5$ , (Figs. 3B and C, 5A and B). The symmetry of this preferred sequence suggests that in the Twist + E12 protein-DNA complex, the Twist and E12 BRs each prefer the same half-site sequence. In contrast, and as observed previously (Blackwell and Weintraub, 1990), MyoD+E12 heterodimers selected a MyoD-like half site at positions +4 and +5, an E2A-like half-site at -4 and -5, and CC or GG bases in the center of the site (Figs. 5A and B), indicating asymmetric binding. Apparently, an E2A BR normally prefers distinct half-

sites in the context of these two bHLH dimerization partners, indicating an intermolecular effect on how it interacts specifically with DNA.

To investigate how heterodimer formation influences the binding preferences of the E12 and MyoD BRs, we performed *in vitro* selection on combinations of MyoD and E12 BR mutants. When the BR of one partner within a MyoD + E12 heterodimer was substituted with that of the other, the heterodimer binding preferences outside the CANNTG consensus corresponded to those of the individual BRs. For example, unlike MD(E12B) homodimers (Figs. 3B and C), heterodimers of MD(E12B)+E12 preferred wild-type heterodimer sequences in the center of the site, and selected E2A-like sequences in both flanking regions, at  $\pm 4$  and  $\pm 5$  (Figs. 5A and B). A heterodimer of MyoD and an E12 protein containing the MyoD BR (E12(MDB); Fig. 2A) similarly selected a wild-type heterodimer preference within the CANNTG motif, but preferred a MyoD-like sequence at  $\pm 4$  and  $\pm 5$  (Figs. 5A and B). In contrast, MD(E12B) + E12(MDB) heterodimers had a binding preference more similar to Twist (Figs. 5A and B), indicating that placement of each BR in the protein context of the other partner affected binding over the entire site. A striking aspect of our findings is that each of the mutant homo- or heterodimer protein complexes that we have examined selected sequences that correspond to particular patterns preferred by MyoD, E2A, or Twist proteins (Figs. 3C and 5B).

These *in vitro* selection findings were supported by assays of binding to individual sites, including a sequence from a muscle-specific regulatory region (MCK-R). This site corresponds to the MyoD + E12 heterodimer *in vitro* binding preference and responds to MyoD *in vivo*, and was used in the original analysis of the myogenic residues (Blackwell and Weintraub, 1990; Davis et al., 1990; Weintraub et al., 1991). In an EMSA, MyoD + E12 heterodimers bound with higher affinity to either the MCK-R or MyoD sites than to the Twist site (Fig. 5C, lanes 3, 12, and 21). MyoD(E12B) + E12 heterodimers only slightly preferred the MCK-R heterodimer site to the Twist site, but appeared to prefer either of these sequences to the MyoD site (Fig. 5C, lanes 5, 14, and 23). As the preferences of MD(E12B-A) and MD(E12B-AT) homodimers would predict, introduction of both A<sub>5</sub> and T<sub>6</sub> into MD(E12B) altered its sequence preferences as a heterodimer with E12, so that they were more similar to those of MyoD (not shown). MyoD + E12(MDB) heterodimers only modestly preferred the MyoD or MCK-R sites in comparison to the Twist site (Fig. 5C, lanes 4, 13, and 22). In contrast, the Twist site was preferred by MD(E12B) + E12(MDB), Twist, and Twist + E12 complexes (Fig. 5C, lanes 6, 8, 9, 15, 17, 18, 24, 26, and 27).

Binding site competition and protein titration assays also supported the *in vitro* selection data. The MyoD site competed more effectively than the Twist site for binding by either MyoD homodimers or MyoD + E12 heterodimers (Figs. 6A and 6B, lanes 1, 4, 7, 10, 13, and 16). In contrast, the Twist site competed more effectively for binding by MD(E12B), MD(E12B) + E12, Twist, and Twist + E12 complexes, although these latter complexes appeared to bind with less specificity than did MyoD + E12 complexes (Figs. 6C and 6D, lanes 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, and 18). However, the distinct binding specificities of MyoD + E12 and Twist + E12 heterodimers were apparent in a protein titration assay, in which the amount of MyoD or Twist protein was varied under conditions of low DNA concentration (Figs. 7A and B, lanes 1-6 and 13-18) that more closely represent differences in binding affinity (Carey, 1991). Also in agreement with results described above (Fig. 5C, lanes 14 and 23), heterodimers of MD(E12B) + E12 bind to the MCK-R site with decreased specificity, and with slightly lower affinity than MyoD + E12 complexes (Figs. 7A and B, lanes 7-12).

To investigate the role of the BR-HLH junction region in BR positioning, we examined the DNA binding preferences of the MD(E12BJ) and E12(MDBJ) mutants, each of which contain both the BR and junction of the other partner (Fig. 2). In contrast to MD(E12B) + E12(MDB) heterodimers (Figs. 5A and B; Fig. 5C, lanes 6, 15, and 24), MD(E12BJ) + E12(MDBJ) heterodimers (Fig. 2A) bound to the MyoD, Twist, and MCK-R sites with relative preferences that are comparable to those of MD + E12 heterodimers (Fig. 5C, lanes 3, 7, 12, 16, 21, and 25). Apparently, the Twist-like sequence preference resulting from simultaneous mis-pairing of both the MyoD and E12 BRs (Fig. 5A and B) can be "corrected" by matching each of these BRs with the corresponding junction region. Similarly, and in contrast to MD(E12B) homodimers, MD(E12BJ) homodimers bind to the MyoD, Twist, and MCK-R sites with preferences that are similar to E2A proteins (Fig. 8B and C, lane 20; not shown). These findings indicate that the BR-HLH junction can be critical for establishing the sequence specificity of an E2A BR, presumably because it influences how the BR is positioned on the DNA.

**Contributions of the BR and junction to binding affinity and specificity.** It has been shown previously that introduction of A<sub>5</sub>, T<sub>6</sub>, and either the junction region or K<sub>15</sub> of MyoD confers upon E12 the capacity to induce myogenesis (Fig. 2)(Davis and Weintraub, 1992). In the MyoD-DNA complex, A<sub>5</sub> and T<sub>6</sub> are not positioned to allow direct protein-protein contact (Fig. 1) (Ma et al., 1994), but we have shown that they are critical for the DNA sequence preferences of MyoD, apparently because they affect the conformation of the BR-DNA complex. We have also determined that the junction region can influence how the E2A BR binds DNA. These observations suggest the possibility that the capacity for myogenesis might derive entirely from the conformation of the DNA-bound MyoD BR, a model which would predict that the respective sequence preferences of each of these bHLH proteins might be established by amino acids at BR positions 5, 6, and 15. We have investigated this model by determining how individual substitutions at these positions, which have been shown to be critical *in vivo*, influence the DNA binding preferences of MyoD.

To address the importance of the MyoD junction region for DNA binding, we substituted MyoD positions 14 and 15 (Fig. 8A), and left position 13 intact because it is not required for the MyoD sequence preference in the MD(E12B-AT) mutant (Figs. 2 and 3C). Substitution of alanine for S<sub>14</sub>, which does not interact with DNA (Ma et al., 1994), increased binding affinity (MD(AK), Fig. 8A; Figs. 8B, and C, lanes 4 and 5), perhaps by stabilizing the BR helix. The preference of MD(AK) for the MyoD site was not substantially altered by replacement of position 15 with alanine (MD(AA)), or with either glutamic acid (MD(AD)) or serine (MD(AS) and MD(QS)), which respectively correspond to residues from E12 and Twist (Fig. 8A; Figs. 8B and C, lanes 5-9). The relative preferences of these mutants for the MyoD site are comparable to the binding preferences of other proteins that were confirmed by binding competition analysis (Figs. 4 and 6). Apparently, appropriately specific DNA binding by MyoD homodimers is not impaired by a variety of BR-HLH junction substitutions, including non-conservative mutations of K<sub>15</sub>. This flexibility contrasts with the importance of the junction region for positioning the E12 BR, and with the requirement for K<sub>15</sub> for myogenesis.

To investigate the role of BR positions 5 and 6 in a neutral context, we first substituted alanine for two non-conserved BR residues (MD-AAATA, Fig. 8A) that are not predicted to be required for DNA binding (Fisher et al., 1993; Ma et al., 1994). This substitution proportionally increased binding to both sites in the context of MyoD (MD-

AAATA; Figs. 8B and C, lane 10), and enhanced specificity for the MyoD site in the context of MD(AA) (Fig. 8A; Figs. 8B and C, lane 12). Replacement of T<sub>6</sub> with asparagine conferred a preference for the Twist site (MD-AAANA, Fig. 8A; Figs. 8B and C, lanes 10 and 13), a finding that parallels the preferences of MD(E12B-AT) and MD(E12B-A) (Figs. 3B and C). This effect was not diminished by various BR-HLH junction mutations or enhanced by presence of Twist junction residues (Figs. 8B and C, lanes 13-17), indicating that N<sub>6</sub> is the most important of these residues for the Twist sequence preference. To test whether E2A amino acids that correspond to the three myogenic residues could specify an E2A-like DNA binding preference, we introduced an asparagine at BR position 7 into MD-AAANA and MD-AAANA(AD), the latter of which contains the D<sub>15</sub> residue characteristic of E2A proteins (Fig. 8A). In contrast to MD(E12BJ), these mutants strongly preferred the Twist site to the MyoD or MCK-R sites (Figs. 8B and C, lanes 18-20; not shown), indicating that establishment of an E2A homodimer sequence preference requires additional E2A BR or junction residues, and that the conformational mechanisms that dictate this asymmetric sequence preference might be complex.

In the examples we have analyzed, MyoD mutants that lack myogenic activity bind preferentially to the Twist site (Figs. 2 and 3C), raising the question of whether changes in DNA binding preferences accompany conversion of E12 into a "myogenic" protein through introduction of MyoD BR and junction residues. E12 homodimers do not bind DNA as well as the E2A protein E47 (Fig. 9, lanes 1, 2, 8, 9, 15, and 16), which also cannot induce myogenesis (Davis and Weintraub, 1992). Introduction of the MyoD BR into E12 is not sufficient for myogenesis (E12(MDB), Fig. 2), but sharply increased binding of E12 to all three sites and was associated with a modest preference for the MyoD site (Fig. 9, lanes 3, 10, and 17). The E12(MDBJ) mutant, which can induce myogenesis (Fig. 2), bound to each of the three sites at a lower level than E12(MDB) and did not have a markedly increased preference for either the MyoD or MCK-R sites (Fig. 9, lanes 4, 11, and 18). Heterodimerization with E47 increased the relative levels with which E12(MDBJ) bound to the MyoD and MCK-R sites (Fig. 9, lanes 6, 7, 13, 14, 20, and 21), but also did not identify DNA binding effects that appear to be sufficient to account for the different functional properties of E12(MDB) and E12(MDBJ). These findings further support the idea that the MyoD junction region is not critical for DNA binding (Figs. 8B and C, lanes 4-9), and instead is important for myogenesis because it is involved in other interactions (Davis and Weintraub, 1992).

## **B. TTP/TIS11 proteins and cell survival:**

### **Programmed cell death in response to TTP.**

To test whether continuous TTP expression might impair cell viability, we introduced TTP into 3T3 cells by transient transfection. Within two days, many TTP-transfected cells appeared apoptotic (Fig. 10A), were positive in a TUNEL assay (Fig. 10B), and contained pyknotic nuclei (Fig. 11B). This cell death increased between 24 and 48 hours after transfection (Fig. 10C), indicating a relatively slow onset.

Within a transfected population, the frequency of cell death was generally proportional to the amount of expression construct introduced, but in individual cells only modest levels were required (Fig. 11). Introduction of either 50 or 200 ng of TTP expression vector triggered apoptosis in a significant fraction of transfected cells (Fig. 11C). In these experiments, transfection efficiencies ranged between 25% and 50% (not shown), but TTP protein could not be detected by western blotting (Fig. 11A, lanes 7

and 8). In contrast, endogenous TTP was readily detectable after serum stimulation (Fig. 11A, lanes 2 and 3); although by immunofluorescence it was barely apparent above background (not shown). After introduction of 50 ng TTP expression vector, most dead cells (72%) did not express TTP at levels that were detectable by immunofluorescence (Fig. 11C). When 200 ng of expression vector was introduced, a moderately higher proportion of apoptotic than non-apoptotic cells expressed TTP at just visible levels (41% versus 25%), but TTP was still undetectable in many apoptotic cells (38%) (Figs. 11B and 11C). In each of these transfections, the proportion of cells expressing TTP at high levels did not appear to be elevated in the apoptotic fraction, perhaps because some of these apoptotic cells had become detached from the plate (not shown). Considering the high transfection efficiency, the observation that a transfected population expressed less TTP than serum-induced cells (Fig. 11A, lanes 2, 3, 7, and 8) suggests that many apoptotic transfected cells expressed TTP at levels that were comparable to or lower than those resulting from serum stimulation. This indicates that excessively high levels of TTP are not required for its induction of cell death.

Various cell types undergo apoptosis in response to TTP, including primary cells (Fig. 10C), demonstrating that immortalization is not a prerequisite. In U2OS and SAOS2 cells, death was delayed (Fig. 10C) and peaked after 72 hours, perhaps because of their slower growth rates (not shown). The lack of functional p53 in SAOS2 cells (Chandar et al., 1992) also may have decreased but did not prevent their apoptotic response, indicating that p53 is not required. The frequency of apoptosis was relatively low in 293 cells, particularly at 48 hours after transfection (Fig. 10C), and was not increased by additional TTP expression vector (not shown). The decreased apoptotic response in this cell line, in which the effects of TTP on the TNF- $\alpha$  mRNA have been studied (Carballo et al., 1998; Lai et al., 1999) is likely to derive from its expression of adenovirus E1B 19K, which mimics the anti-apoptotic protein Bcl-2 (Han et al., 1998). The *C. elegans* POS-1 and PIE-1 proteins each contain two related zinc fingers (Mello et al., 1996; Tabara et al., 1999) but did not cause significant apoptosis in this assay (not shown), indicating that not all Cys<sub>3</sub>His zinc finger proteins trigger cell death when constitutively expressed. Cell death in response to TTP was decreased by mutation of the first Cys<sub>3</sub>His zinc finger, and abrogated by alteration of both (Fig. 12), supporting the idea that it is caused by TTP acting on appropriate targets.

### **Similarities between TTP/TIS11- and oncogene-induced apoptosis.**

When expressed constitutively, TIS11b and TIS11d also triggered apoptosis associated with TUNEL-positive nuclei (not shown). Each TTP/TIS11 protein induced cell death with similar frequency and timing over a range of expression vector amounts (Figs. 13A and B), and also as fusions with green fluorescent protein (GFP; not shown). We have not developed assays for detecting endogenous TIS11b and TIS11d proteins, nor could we detect the corresponding TTP/TIS11-GFP fusion proteins after introduction of the modest DNA amounts used in Figures 13A and B. When these GFP fusion proteins were overexpressed as in Figures 11A (lane 9) or 12B, however, western blotting with a GFP antibody revealed that they were present at similar levels (not shown). This suggests that the three TTP/TIS11 proteins are also similarly expressed when introduced at low levels in cell death assays, and that they trigger apoptosis comparably.

The hypothesis that TTP/TIS11 proteins affect cell growth or survival signals suggests that they might stimulate apoptosis analogously to some oncoproteins. The immediate early protein c-Myc is involved in G1 entry, and possibly in physiological

apoptotic events (Shi et al., 1992; Zhan et al., 1997). Its forced expression under low serum conditions triggers apoptosis over a similar time course as TTP/TIS11 proteins, apparently by stimulating growth or apoptotic pathways in the absence of survival signals (Juin et al., 1999). This apoptosis is enhanced when c-Myc is overexpressed, but can be detected when it is expressed constitutively at the levels that are observed following serum induction (Evan et al., 1992). It is inhibited when survival signals are restored by treatment with insulin-like growth factor 1 (IGF-1), it requires the mitochondrial death machinery, and it can involve p53 and cell-surface death receptors (Juin et al., 1999). In apparent contrast, the S-phase transcription factor E2F-1 induces apoptosis in the presence of serum, and can do so independently of p53 when it is expressed at sufficient levels (Hsieh et al., 1997; Kowalik et al., 1995; Phillips et al., 1997).

Cell death in response to TTP/TIS11 proteins appears to be analogous to these apoptotic events in various respects. In each case, it was prevented by co-expression of Bcl-2, which inhibits the mitochondrial death machinery but not direct caspase activation by death receptors (Fig. 13C)(Gross et al., 1999). It was also partially alleviated by the CrmA protein (Fig. 13C), an effective inhibitor of death receptor-activated caspases (Ashkenazi and Dixit, 1998). The latter finding suggests that TTP/TIS11-stimulated apoptosis might involve death receptors, but it is also possible that its inhibition by CrmA derives from effects on caspases activated by the mitochondrial machinery (Gross et al., 1999). Serum withdrawal is not a precondition for TTP/TIS11-stimulated apoptosis, but it markedly increased the frequency of death (Fig. 13D). This increase was not abrogated by nutrient replacement, and was offset by IGF-1 treatment (Fig. 13D), supporting the model that it involves a lack of survival signals. Apparently, constant TTP/TIS11 protein expression can overcome or circumvent the survival signals provided by serum, but also stimulates apoptosis more rapidly when these signals are lacking.

### **Synergistic induction of apoptosis by TTP and TNF- $\alpha$ .**

TNF- $\alpha$  stimulates apoptosis by binding to its Type I receptor, an event which triggers caspases directly (Ashkenazi and Dixit, 1998). Simultaneously, however, this binding can activate anti-apoptotic genes, which apparently must remain silent if cell death is to occur (Grumont et al., 1999; Wang et al., 1998; Wu et al., 1998; Zong et al., 1999). TNF- $\alpha$ -induced apoptosis is enhanced by expression of the oncoproteins c-Myc, adenovirus E1A, and E2F-1, and by lack of the tumor suppressor Rb, and it is impaired by inhibition of either c-Myc or cell cycle progression (Janicke et al., 1994; Klefstrom et al., 1994; Meikrantz and Schlegel, 1996; Phillips et al., 1999). These findings indicate that pathways which regulate growth or proliferation can influence how a cell responds to TNF- $\alpha$ .

TNF- $\alpha$  treatment induces TTP mRNA expression (Carballo et al., 1998), suggesting the possibility that TTP/TIS11 proteins might also affect responses to TNF- $\alpha$ . To test this idea, we added TNF- $\alpha$  to 3T3 cells that were transfected with TTP/TIS11 expression vectors in amounts that triggered only modest cell death (Fig. 14). Administration of TNF- $\alpha$  shortly after transfection dramatically increased apoptosis in cells that expressed TTP but, surprisingly, not TIS11b or TIS11d (Fig. 14A). The same trends were observed when TNF- $\alpha$  was added after TTP had been expressed for 24 hours after transfection (Fig. 14B). In these TTP-expressing cells, death was also increased by incubation with TNF- $\alpha$  for only four hours in the presence of

cycloheximide (Fig. 14C), which blocks its induction of anti-apoptotic proteins. The rapidity of this last effect is more consistent with apoptosis induced by death receptors (Ashkenazi and Dixit, 1998) than with the slower time course of TTP-stimulated cell death (Fig. 10C), suggesting that TTP has sensitized these cells to the apoptotic stimulus of TNF- $\alpha$ . Similar results were obtained in parallel experiments performed in HeLa cells (not shown). Supporting the idea that this effect is specific to TTP, expression of TIS11b and TIS11d at higher levels increased the overall level of cell death, but did not promote induction of additional apoptosis by TNF- $\alpha$  (not shown).

### **KEY RESEARCH ACCOMPLISHMENTS**

-- Demonstration that the conformation with which bHLH proteins bind DNA is of critical importance, and that binding affinity *per se* is not necessarily sufficient for their activity. These observations may have important implications for understanding functions of c-Myc, which have been implicated in breast cancer.

-- The immediate early protein TTP, which is induced by EGF and various mitogens, causes apoptosis when it is expressed constitutively at levels to which it is normally induced transiently. This apoptosis appears analogous to that induced by constitutive expression of c-Myc, E2F-1 and other oncogenes, and appears to involve affects on growth and survival pathways. The data indicate that TTP and the related TIS11 proteins act on these pathways when they are expressed transiently.

-- TTP in particular sensitizes cells to TNF- $\alpha$ -induced apoptosis, perhaps similarly to c-Myc and E2F-1. This finding could be relevant to the phenotype of TTP knockout mice, which suffer from a widespread polyinflammatory and myeloproliferative syndrome that is mediated by TNF- $\alpha$ .

### **REPORTABLE OUTCOMES**

#### **Manuscripts:**

Huang J, **Blackwell TK**, Kedes L, Weintraub H. Differences between MyoD DNA binding and activation site requirements revealed by a functional random sequence selection. *Mol Cell Biol* 1996;16:3893-00.

**Blackwell TK**, Kophengnavong T, Johnson BA. Biochemical and functional analyses of c-Myc- and TTP-related proteins. Abstract for 1997 Era of Hope meeting.

De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, **Blackwell TK**, Blackshear PJ. Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression. *Gene* 1999;228:133-45.

Kophengnavong T, Michnowicz JE, **Blackwell TK**. Establishment of Distinct MyoD, E2A, and Twist DNA binding specificities by different basic region-DNA conformations. *Mol Cell Biol* 2000; 20: 261-272.

Johnson B, Geha M, **Blackwell TK**. Similar but Distinct Effects of the Tristetraprolin/TIS11 Immediate-Early Proteins on Cell Survival. *Oncogene* 2000, pending minor revisions.

#### Funding:

Molecular analyses of TTP, a key regulator of TNF-alpha activity. B. A. Johnson, PI. Postdoctoral fellowship from the Arthritis Foundation. 7/1/99-6/30/01.

Molecular analyses of TTP/TIS11 protein functions. T. K. Blackwell, PI. NIH-RO1 in revision for 3/1/00.

### CONCLUSIONS

#### A. Determinants of bHLH protein DNA binding specificity:

**bHLH protein DNA binding specificity deriving from effects on BR-DNA conformation.** The myogenic MyoD BR residues A<sub>5</sub> and T<sub>6</sub> are essential for myogenesis, but not for binding of MyoD + E2A heterodimers to a muscle-specific site *in vitro* or *in vivo* (Davis and Weintraub, 1992; Weintraub et al., 1991). However, we have determined that these residues are required for MyoD to bind DNA with its characteristic specificity for particular CANNTG sites. Substitution of asparagine for T<sub>6</sub>, and especially for both A<sub>5</sub> and T<sub>6</sub>, results in MyoD binding preferentially to a Twist site (Figs. 8B and C, lanes 10, 13, and 18). The Twist-like MD(E12B) sequence preference is affected partially by substitution of A<sub>5</sub> for the corresponding asparagine (MD(E12B-A), Fig. 3C), but is reconfigured by introduction of both A<sub>5</sub> and T<sub>6</sub> so that it is indistinguishable from that of wild-type MyoD (MD(E12B-AT), Fig. 3C). The data indicate that MyoD residues A<sub>5</sub> and T<sub>6</sub> are each critical for its DNA binding sequence preferences, and that the N<sub>6</sub> residue, which is common to the Twist and MD(E12B-A) BRs (Fig. 2), is important for the Twist-like preference. Mutations of these individual BR residues alter sequence preferences across each half-site (Fig. 3C), raising the question of how they might have such a global effect on how the BR helices and the DNA interact preferentially with each other.

A structure of MyoD obtained by X-ray crystallography suggests how A<sub>5</sub> and T<sub>6</sub> might influence binding sequence specificity. When bound to its preferred recognition site, MyoD does not directly contact base pairs that it specifies in the center of and flanking the CANNTG consensus (Ma et al., 1994). However, A<sub>5</sub> and T<sub>6</sub> allow the MyoD BR helix to pack more tightly into the major groove than do the corresponding N<sub>5</sub> and N<sub>6</sub> residues of E2A proteins, in part because of their smaller sizes (Figs. 1 and 2)(Ma et al., 1994). As a result, the MyoD BR residues T<sub>6</sub> and R<sub>2</sub> directly contact CANNTG bases at  $\pm 2$  and  $\pm 3$  respectively, and R<sub>1</sub> binds a backbone phosphate at  $\pm 6$  (Fig. 1)(Ma et al., 1994). In contrast, in E47 R<sub>2</sub> swings out of the major groove and contacts the backbone, and the residue at position 1 does not interact directly with the DNA (Brownlie et al., 1997; Ellenberger et al., 1994). Supporting the idea that A<sub>5</sub> and T<sub>6</sub> influence the conformation of the DNA-bound BR, substitution of asparagine for A<sub>5</sub> in MyoD increases its sensitivity to protease digestion (Huang et al., 1998). Our findings suggest that protein-DNA interactions that depend specifically upon the MyoD A<sub>5</sub> and



T<sub>6</sub> residues may directly influence how the BR helix interacts preferentially with the DNA, and thereby indirectly specify its characteristic sequence preferences at positions within and flanking the CANNTG consensus.

Such indirect conformational effects also appear to be critical for the E2A and Twist sequence preferences. When E47 homodimers bind DNA, a single subunit contacts a base in the center of the site through R<sub>10</sub> (Fig. 2). This interaction could be important for the asymmetric E2A homodimer sequence preference (Ellenberger et al., 1994). However, the Twist-like sequence preference that is characteristic of Twist + E2A heterodimers and MD(E12B) homodimers is different across each 5 bp half-site and symmetric (Figs. 3C and 5B), suggesting that it is likely to be established indirectly, through an intermolecular effect that involves a distinct positioning of the BR helix. Introduction of the E12 BR-HLH junction region into MD(E12B) "corrects" its binding preference so it is like that of E2A homodimers (MD(E12BJ), Fig. 5C, lanes 7, 16, and 25; Figs. 8B and C, lane 20), implicating the BR-HLH junction in this effect. Presumably, the E2A junction acts in concert with the asparagines at BR positions 5 and 6 (Fig. 2), although the Twist-like preference of the MD-AANNA(AD) mutant (Figs. 8B and C, lane 19; not shown) suggests that the E2A junction residue D<sub>15</sub> is not sufficient. The finding that E2A proteins can be targeted to different DNA sequences by different dimer partners may have important implications for their *in vivo* functions.

In contrast, the BR-HLH junction region does not have a strong influence on the MyoD DNA binding preference. Various MyoD junction mutations do not substantially diminish its preference for a MyoD site (Figs. 8B and C, lanes 5-9). In addition, the similar sequence preferences of E12(MDB) and E12(MDBJ) homodimers (Fig. 9, lanes 3, 4, 10, 11, 17, and 18) contrast sharply with the different specificities of MD(E12B) and MD(E12BJ) (Figs. 3D, lanes 2 and 6, and 8B and C, lane 20). This apparent difference between MyoD and E2A proteins might derive from the distinct arrangement of the BR helix on the DNA that results from presence of MyoD residues A<sub>5</sub> and T<sub>6</sub>.

It is striking that, as a group, these various bHLH mutants and dimer combinations bind DNA with a limited number of discrete sequence preferences (Figs. 3C and 5B). Presumably, each of these preferences reflects a preferred conformational "state" that is dictated by how each BR helix and the corresponding DNA sequence conform to each other in an induced fit (Spolar and Record Jr., 1994). This mechanism for recognizing particular CANNTG sites appears to be different from the direct recognition of central bases that is characteristic of bHLH proteins that contain R<sub>13</sub>, and bind to CACGTG or CATGTG sites (Ferre-D' Amare et al., 1993; Ferre-D' Amare et al., 1994; Shimizu et al., 1997). Consistent with this idea, BR residues 5 and 6 do not appear to be important for the function of the R<sub>13</sub>-containing bHLH protein c-Myc (Bodis et al., 1997). In E2A and its tissue-specific dimerization partners, a more flexible conformation-based mechanism might have evolved to increase adaptability in both sequence recognition and function, so that different combinations of these proteins can result in distinct protein-DNA conformations that correspond to particular DNA sequence preferences. Such a model may be particularly plausible for bHLH proteins, because folding of the BR into an  $\alpha$ -helix is driven by its interaction with the DNA (Anthony-Cahill et al., 1992).

**BR-DNA conformation, DNA binding specificity, and myogenesis.** The observation that the MyoD junction and K<sub>15</sub> are not required for an appropriate DNA binding specificity (Fig. 8B and C, lanes 6-9; Fig. 9), supports the model that K<sub>15</sub> is involved in other essential interactions (Davis and Weintraub, 1992). However, our experiments

also pose the question of how the functional importance of A<sub>5</sub> and T<sub>6</sub> might be related to their effects on DNA recognition. Of the MyoD BR mutants we have analyzed, those that do not induce myogenesis bind to DNA as homodimers with a Twist-like preference (MD(E12B) and MD(E12B-A), Figs. 2 and 3C). Heterodimers of MD(E12B) with E12 prefer a heterodimer site (Fig. 5B), but with markedly diminished specificity compared to MyoD + E12 dimers (Fig. 5C, lanes 3, 5, 12, 14, 21, and 23; Fig. 6; Fig. 7A and B, lanes 1-12). This suggests that, at least in part, A<sub>5</sub> and T<sub>6</sub> might be significant for myogenesis because they restrict the DNA binding specificity of MyoD and other myogenic bHLH proteins, so that they are less likely to bind inappropriate sites. However, other observations support a role for the A<sub>5</sub> and T<sub>6</sub> residues in protein-protein interactions. They have been implicated in binding to other proteins off the DNA (Hamamori et al., 1997; Molkentin et al., 1995), and evidence indicates that they are required for activation domain exposure (Black et al., 1998; Huang et al., 1998; Weintraub et al., 1991), and cooperative DNA binding (Bengal et al., 1994). Finally, unlike MyoD, MD(E12B) can activate transcription of a reporter only in particular cell lines, implicating the BR in protein-protein interactions (Weintraub et al., 1991).

In light of evidence that A<sub>5</sub> and T<sub>6</sub> establish the conformation of the DNA-bound BR, it is an attractive model that this effect might influence the function of myogenic bHLH proteins directly, by affecting their interactions with other proteins. Given that relatively subtle alterations of the MyoD BR and junction region can enhance MyoD DNA binding significantly (MD(AK) and MD(AAATA), Fig. 8B and C, lanes 4, 5, and 10), it appears likely that cooperative protein-protein interactions with the BR and junction could influence binding affinity. It has been demonstrated recently that MyoD binds cooperatively with other DNA binding proteins to a particular muscle-specific promoter (Biesiada et al., 1999). The E box sequences through which MyoD activates transcription in the context of this promoter can differ from those it binds preferentially *in vitro* (Huang et al., 1996), suggesting that DNA sequence recognition may be influenced by interactions with cooperating proteins *in vivo*. In addition, interactions with cooperating proteins might be influenced in turn by the specificity of DNA sequence recognition, as suggested by evidence that for MyoD and E proteins the choice between homo- or heterodimer formation may be dictated by the DNA binding affinities of the individual BRs (Maleki et al., 1997; Wendt et al., 1998). Our findings are consistent with the idea that deceptively subtle aspects of sequence recognition could be important for the biological activity of MyoD, if they influence functionally critical interactions that might also involve K<sub>15</sub>, or other MyoD regions. They also suggest that manipulation of these interactions could have important consequences for the functions of these proteins, and that these concepts may be applicable to DNA recognition by other bHLH proteins such as c-Myc.

## **B. TTP/TIS11 proteins and cell survival:**

When expressed constitutively at approximately physiological levels, the TTP/TIS11 immediate early proteins each trigger apoptosis with comparable frequency and timing (Figs. 13A and B), suggesting that they act similarly to each other on related or overlapping regulatory pathways. Various findings suggest that they induce cell death analogously to oncoproteins such as c-Myc, E1A, and E2F-1 (Han et al., 1998; Hsieh et al., 1997; Juin et al., 1999; Kowalik et al., 1995; Phillips et al., 1997; Phillips et al., 1999). For example, in contrast to the rapid stimulus associated with death receptor triggering (Fig. 14C), TTP/TIS11 proteins cause apoptosis over 24-to-48 hours (Figs.

13A and B). This apoptosis is accelerated when survival signals are absent (Fig. 13D) and is dependent upon the mitochondrial machinery (Fig. 13C), which is responsive to abnormal growth regulation, survival signals, and stress (Gross et al., 1999). TTP in particular also appears to be analogous to these oncoproteins in that it can sensitize cells to induction of apoptosis by TNF- $\alpha$  (Fig. 14). Unlike c-Myc, however, TTP/TIS11 proteins can trigger apoptosis in the presence of serum, and in this respect may be comparable to E2F-1, which stimulates DNA replication and cell cycle progression (Nevins, 1998; Phillips et al., 1999). In some experiments, TTP expression increased the number of cells present at 24 hours after transfection (not shown), suggesting that it might also promote proliferation.

Precedents set by various other Cys<sub>2</sub>His zinc finger proteins predict that TTP/TIS11 proteins are likely to have RNA-associated functions (Barabino et al., 1997; Guedes and Priess, 1997; Murray et al., 1997; Rudner et al., 1998; Tabara et al., 1999; Tronchere et al., 1997), and evidence indicates that TTP can bind and influence the stability of TNF- $\alpha$  and other ARE-containing cytokine mRNAs (Carballo et al., 1998; Lai et al., 1999). We have determined that TTP/TIS11 proteins influence cell growth and survival mechanisms, suggesting that they may also act on mRNAs that are involved in those pathways. During growth factor responses, the localization, stability, and translation of various mRNAs are regulated through AREs that are distinct from but related to those of cytokine mRNAs (Chen and Shyu, 1995; Ross, 1995). Such mRNAs might be candidate TTP/TIS11 protein targets, but it will be important to discriminate among the direct and indirect effects of these proteins because an extremely large number of genes are regulated during these responses (Iyer et al., 1999).

The apoptotic effect of expressing TTP/TIS11 proteins constitutively might derive simply from their being present at inappropriate phases of the cell cycle. However, other immediate-early proteins that are associated with growth or proliferation are involved in apoptosis (Hafezi et al., 1997; Shi et al., 1992; Zhan et al., 1997), and TTP is expressed during apoptotic events (Harkin et al., 1999; Mesner et al., 1995), suggesting that TTP/TIS11 proteins might normally promote apoptosis in some contexts. The observation that TTP can sensitize cells to the apoptotic effects of TNF- $\alpha$  is consistent with this idea. The apparently similar sensitization by c-Myc has been proposed to involve the mitochondrial death machinery (Jain et al., 1999), and the analogous effects of E2F-1 have been linked to down-regulation of anti-apoptotic mechanisms (Phillips et al., 1999). TIS11b and TIS11d can stimulate the mitochondrial death machinery (Fig. 13C) but do not sensitize cells to TNF- $\alpha$ -induced apoptosis (Fig. 14), suggesting that TTP acts on additional pathways. TTP might influence TNF- $\alpha$ -induced apoptosis by acting on growth related pathways distinct from those affected by TIS11b and TIS11d, or by interfering with anti-apoptotic gene expression.

Our findings raise the question of why, in mice, lack of TTP causes a specific defect in TNF- $\alpha$  regulation. Although changes in overall TIS11b and TIS11d mRNA levels have not been detected in various TTP<sup>-/-</sup> mouse tissues (Taylor et al., 1996), it remains possible that TTP/TIS11 proteins could have partially redundant functions. Alternatively, TTP might function in growth regulatory pathways that are largely redundant with other mechanisms. TNF- $\alpha$  expression is regulated by a complex array of transcriptional and post-transcriptional mechanisms that respond to numerous inputs (Beutler et al., 1992). By suggesting that TTP expression influences cell growth or survival pathways, and sensitizes cells to TNF- $\alpha$  induced apoptosis, our experiments have identified avenues through which it might influence TNF- $\alpha$  expression or responses indirectly. They indicate that elucidation of how TTP/TIS11 proteins act on

their targets not only will reveal critical aspects of TNF- $\alpha$  regulation, but also may uncover post-transcriptional gene regulation mechanisms that are involved in cellular responses to multiple stimuli, including EGF and various mitogens.

### **C. Figure legends:**

#### **Figure 1.**

A MyoD-DNA complex. In this X-ray crystallographic structure (Ma et al., 1994), a MyoD homodimer is bound to the sequence AACAGCTIGTT, which corresponds to its preferred recognition consensus (Blackwell and Weintraub, 1990). Residues are numbered as in full length MyoD, and their positions as specified as in Fig. 2 and the text are indicated in parentheses. Binding site positions  $\pm 5$  (numbered as in Fig. 3A) are indicated by grey numerals. Side chains are shown only for the myogenic residues (green) (Davis and Weintraub, 1992) and Arg 111 (R<sub>2</sub>) (gold).

#### **Figure 2.**

Myogenic activity of MyoD and E12 BR and junction mutants. Each of these mutants has been described previously (Davis and Weintraub, 1992; Weintraub et al., 1991), and is compared with sequences from mouse MyoD, E12, and Twist. Amino acids that are identical to those of MyoD are underlined, positions that are conserved in most bHLH proteins are shaded, and entire BR and junction regions that have been swapped are bracketed. The column "muscle" indicates the relative activity of these proteins when assayed previously by transfection for conversion of cultured cells into muscle (Davis and Weintraub, 1992; Weintraub et al., 1991). The "++++" indicates the frequency of myogenic conversion obtained with wild-type MyoD, the "++" 30-50% of that obtained with MyoD, and the "+" indicates 5-30% of that obtained with wild-type MyoD. "No" indicates that myogenic conversion was not detected, and "ND" indicates not done.

#### **Figure 3.**

*In vitro* selection assay of binding site preferences. (A) Core sequences of the random sequence oligonucleotide libraries D3 and D6 (Blackwell et al., 1990; Blackwell and Weintraub, 1990). In each library, the bases shown are flanked by sequences which correspond to primers (A and B) that allow selected sequences to be recovered by PCR. A' indicates that primer A corresponds to the opposite strand. (B) Sequences of preferred binding sites. Starting with the D6 oligonucleotide random sequence library (A), three rounds of sequential selection and PCR amplification were performed for binding to the proteins indicated. A sample of the final selected population of binding sites was then sequenced directly as a pool and analyzed by autoradiography. The MyoD preferences at positions  $\pm 1$  described previously (Blackwell and Weintraub, 1990) are more prominent after additional selection rounds (not shown). (C) Summary of sequence preferences identified by *in vitro* selection in (B). MyoD and E2A homodimer preferences were described in (Blackwell and Weintraub, 1990). Binding site positions are numbered as in B, and grey letters indicate bases that were selected against. The CANNTG consensus that was fixed in these experiments is underlined. (D) Binding of MyoD BR mutants to individual oligonucleotide sites, which differed only at the sequences shown. In this EMSA, which was analyzed by phosphorimaging, each sample contained the indicated *in vitro* translated protein at a concentration of 40 pM,

and DNA that was labeled to the same specific activity at 550 pM. Specific and background species are indicated by open and closed triangles, respectively.

#### **Figure 4.**

Specificity of MyoD BR mutant DNA binding. (A) Competition analysis of binding to the labeled MyoD preferred site, analyzed by EMSA and autoradiography. The indicated *in vitro* translated proteins and DNA labeled to the same specific activity were present at concentrations of 50 pM and 900 pM, respectively. When the samples were mixed, unlabeled competitor DNA sites were added at the indicated ratios relative to the labeled probe. (B) Competition analysis of binding to the Twist preferred site, performed as in (A).

#### **Figure 5.**

Binding site preferences of MyoD, E2A, and Twist heterodimer complexes. (A) *In vitro* selection analysis of binding site preferences. Four rounds of selection from the D3 library (Fig. 3A) were performed as for each *in vitro* translated protein complex. In each case, the heterodimer complex could be easily identified in the EMSA on the basis of mobility (Blackwell and Weintraub, 1990), particularly because E12 homodimers bind DNA poorly (Fig. 9). In the Twist homodimer selection, binding to Twist/E12 heterodimers was selected for in the first round, because of the relatively low level of Twist homodimer binding. Subsequent rounds were performed using Twist homodimers. Each sample was analyzed by sequencing and autoradiography as in Fig. 3B. (B) Summary of sequence preferences identified in (A), depicted as in Fig. 3C. MyoD + E2A heterodimer preferences were also described previously in (Blackwell and Weintraub, 1990). (C). Binding of bHLH heterodimers to individual preferred sites, analyzed by EMSA and phosphorimaging. E2A-derived proteins were present at a concentration of 8 pM, and Twist and MyoD-derived proteins at 19 pM. The indicated DNA sites that had been labeled to the same specific activity were present at 550 pM. The MCK-R site differs from the others only at the positions shown. A background species is indicated by a closed triangle.

#### **Figure 6.**

Binding competition analysis of DNA binding by bHLH heterodimers. In (A) and (B), binding of the indicated protein complexes to the labeled MyoD site (Fig. 3D) was competed by addition of an unlabeled binding site at ratios indicated above the gel. These experiments were performed and analyzed as in Fig. 4, except that labeled DNA was present at 600 pM, E12 protein present at 8 pM, and all other proteins at 19 pM. In (C) and (D), binding of the indicated protein complexes to the labeled Twist site (Fig. 3D) was competed by addition of the indicated unlabeled sites. These experiments were performed in (A) and (B), except that labeled DNA was present at 1.1 nM, and they were analyzed by autoradiography. Note that the gel shown in C was exposed longer than that shown in D, as indicated by comparison of lanes 1-6. A background species is indicated by a closed triangle.

#### **Figure 7.**

Protein titration of DNA binding by bHLH heterodimers. (A) Binding to the Twist site, analyzed by EMSA and phosphorimaging. In each experiment, E12 was present at 8 pM and DNA that had been labeled to the same specific activity at 5 pM.

The indicated partner proteins were present at the concentrations shown above the gel (in pM). (B) Binding to the MCK-R site, analyzed as in (A).

### **Figure 8.**

Effects of bHLH BR and BR-HLH junction residues on MyoD binding preferences. (A) Mutagenesis analysis of the MyoD BR and junction. MyoD BR mutant sequences are compared with the MyoD, E12, and Twist BRs (Fig. 2). Conserved bHLH residues are shaded, and residues that are altered within full-length MyoD are underlined. (B) Binding of MyoD mutants described in (A) to the MyoD preferred site. These mutants are compared with the indicated wild-type proteins, and binding is assayed as in Fig. 3D, except that each protein is present at 40 pM, and DNA labeled to the same specific activity present at 400 pM. E47 is an alternatively-spliced E2A protein that binds DNA well as a homodimer (Murre et al., 1989). (C) Binding of MyoD mutants to the Twist preferred site, assayed as in (B).

### **Figure 9.**

DNA binding by E12 mutants. DNA binding by the indicated protein complexes is assayed as in Fig. 5C, except that all E12 derivatives are present at 8 pM, and E47 at 19 pM. A protein-DNA complex of intermediate mobility that corresponds to E47-E12 heterodimers is indicated by an asterisk, and a background species by a closed triangle.

### **Figure 10**

Cell death in response to TTP. (A) Blue cell assay of TTP-induced death. Cells were transfected with 1.5  $\mu$ g expression plasmid (either empty vector CS2 or CS2TTP), and 0.5  $\mu$ g  $\beta$ -gal reporter plasmid, then X-gal stained after 48 hours and shown at 40X magnification. (B) TUNEL assay. Cells were stained for TUNEL activity at 24 hours after lipofectamine plus transfection (with either 1  $\mu$ g CS2TTP or CS2 vector control expression plasmid), or after treatment with TNF- $\alpha$  (100ng/ml) and cycloheximide (30 $\mu$ g/ml) for 3 hours as a positive control. Typical fields are shown. No staining was detected in a parallel experiment that lacked the TUNEL reagent (not shown). (C) TTP-induced cell death in cell lines and primary cells (designated by an asterisk). In a similar experiment to (A), cells were transfected with 200 ng of CS2 vector or CS2TTP, and 100 ng of  $\beta$ -gal reporter plasmid. After 24 or 48 hours they were X-gal stained and the percentage of dead blue cells was determined. Numbers indicate the mean of four transfections  $\pm$  the standard deviation. Human foreskin keratinocytes are indicated by HFK, and mouse embryo fibroblasts by MEF.

### **Figure 11**

Cell death caused by forced expression of TTP at modest levels. (A) Expression of TTP after serum induction or transfection of TTP expression plasmid, assayed by western blotting with affinity purified TTP antibody. 3T3 cells were serum stimulated (Taylor et al., 1996), or transfected with the indicated amount of TTP expression vector. Each lane contained 200  $\mu$ g total protein. The 43kD species present in lanes 2, 3, and 9 corresponds to TTP. A background band similar to that found in all lanes in this gel (indicated by an asterisk) has been detected by a different antiserum against the same peptide (Carballo et al., 1998). The more slowly-migrating TTP-specific bands (lanes 9 and 10) can be converted to faster-migrating species by phosphatase treatment (lane 11), suggesting that they represent phosphorylated TTP forms described previously

(Taylor et al., 1995). The western blot shown in lanes 10 and 11 was performed using TTP antiserum that had not been affinity purified, and does not detect the background species apparent in lanes 1-9. (B) TTP expression in transfected cells. Typical fields are shown. 24 hours after transfection with the indicated amounts of TTP vector and either 100 or 200 ng pN3eGFP, cells were fixed and stained with affinity-purified TTP antibody (1 $\mu$ g/ml), which was detected with a Cy3-conjugated secondary antibody. Nuclei are revealed by Hoechst staining, and transfected cells by GFP autofluorescence. Additional GFP-positive cells with low fluorescence levels were visible under the microscope. A dual image shows the overlap between GFP fluorescence and TTP staining. Arrows on Hoechst-stained fields indicate apoptotic cells in which TTP staining is indistinguishable from background, and which are labeled in the TTP field with **u** (undetectable). Cells that have high TTP staining levels are labeled with **h**, and that have just visible TTP with **v**. (C) TTP staining levels compared with apoptosis. Cells were transfected with the indicated amounts of TTP vector, then 24 hr. later the percentage of cell death was determined as in Fig. 10C. In a control transfection lacking TTP expression vector, 5% of the cells were apoptotic, all of which were GFP-negative (not shown). In a parallel experiment, cells were transfected on coverslips with the same amounts of TTP expression vector and the GFP reporter plasmid as in B. After staining, TTP expression was scored as in B (**u**, **v**, or **h**) in approximately 50 apoptotic and 300 non-apoptotic GFP positive cells. The percentages indicated in the table refer to the proportion of each group (apoptotic or non-apoptotic) within each staining category.

### Figure 12

The TTP zinc fingers are required for apoptosis. (A) The M1 and M1,2 mutants, in which zinc finger residues that were substituted within full-length TTP are highlighted. (B) TTP mutant expression, assayed by western blotting following transfection of 2  $\mu$ g of the indicated plasmids. Each lane of the gel, which was not run as far as that shown in Fig. 11A, contained 100  $\mu$ g total protein. TTP was detected using serum that had not been affinity purified (Fig. 11A). (C) Cell death caused by TTP mutants. Cells transfected with 200 ng of the indicated constructs were assayed for cell death 48 hours after transfection as in Fig. 10C.

### Figure 13

Cell death in response to each TTP/TIS11 protein. In (A) and (B), 3T3 cells were transfected with the indicated amounts of TTP, TIS11b, TIS11d, or CS2 control vectors along with 100 ng of  $\beta$ -gal reporter, then cell death was assayed as in Fig. 10C after either 24 (A), or 48 hours (B). (C) Suppression of TTP/TIS11-induced apoptosis by Bcl-2 and CrmA. 200 ng of the indicated expression vector and 100ng of  $\beta$ -gal reporter were co-transfected into HeLa cells. For Bcl-2 inhibition (white bars), 200 ng of CMV-Bcl-2 were added, and for CrmA inhibition (shaded bars), 1.7  $\mu$ g of CMV-CrmA were added. After 24 hours, death was assayed as in Fig.10C. (D) Enhancement of TTP/TIS11-induced apoptosis by serum deprivation. 3T3 cells were lipofectamine plus transfected with 100ng of effector plasmid as in (C). After 3 hours, they were incubated in medium containing either 0.1% serum (dark gray bars), 0.1% serum plus 100 ng/ml IGF-1 (Sigma; hatched bars), 0.1% serum plus serum replacement (SR) medium 1 (Sigma; mid-gray bars), 0.1% serum plus IGF-1 and SR1 (white bars) or 10% serum (black bars). SR1

medium provides nutrient replacement but not growth factors. Cell death was assayed 21 hours later.

### **Figure 14**

Synergistic induction of cell death by TTP and TNF- $\alpha$ . (A) Effects of TNF- $\alpha$  treatment shortly after transfection. 3T3 cells were transfected with 25 ng of the indicated expression vector and 100 ng of  $\beta$ -gal reporter, using lipofectamine plus (Gibco BRL). Recombinant mouse TNF- $\alpha$  (R & D Systems) was added to the indicated concentration after 3 hours, and 19 hours later cell death was assayed as in Fig. 10C. (B) Addition of TNF- $\alpha$  after 24 hours of TTP expression. This experiment was conducted as in (A), except that TNF- $\alpha$  was added 24 hours after transfection, and cell death assayed 24 hours later. (C) Addition of TNF- $\alpha$  and cycloheximide after 24 hours of TTP expression. This experiment was conducted as in B, except that cycloheximide was added to 10 $\mu$ g/ml along with TNF- $\alpha$ , and cell death was assayed 4 hours later.

### **REFERENCES**

- Aksan, I., and Goding, C. R. (1998). Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-Box elements *in vitro* and *in vivo*. *Mol Cell Biol* 18, 6930-8.
- Anthony-Cahill, S. J., Benfield, P. A., Fairman, R., Wasserman, Z. R., Brenner, S. L., Stafford, W. F., Altenbach, C., Hubbell, W. L., and DeGrado, W. F. (1992). Molecular characterization of helix-loop-helix peptides. *Science* 255, 979-983.
- Ashkenazi, A., and Dixit, V. M. (1998). Death receptors: signaling and modulation. *Science* 281, 1305-1308.
- Barabino, S. M. L. X., Hübner, W., Jenny, A., Miniville-Sebastia, L., and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev.* 11, 1703-1716.
- Batchelder, C., Dunn, M. A., Choy, B., Suh, Y., Cassie, C., Shim, E. Y., Shin, T. H., Mello, C., Seydoux, G., and Blackwell, T. K. (1999). Transcriptional repression by the *C. elegans* germline protein PIE-1. *Genes Dev.* 13, 202-212.
- Bengal, E., Flores, O., Rangarajan, P. N., Chen, A., Weintraub, H., and Verma, I. M. (1994). Positive control mutations in the MyoD basic region fail to show cooperative DNA binding and transcriptional activation *in vitro*. *Proc Natl Acad Sci U S A* 91, 6221-5.
- Beutler, B., Han, J., Kruys, B., and Giroir, B. P. (1992). Coordinate regulation of TNF biosynthesis at the levels of transcription and translation. Patterns of TNF expression *in vivo*. In *Tumor necrosis factors: the molecules and their emerging role in medicine*, B. Beutler, ed. (New York, NY: Raven Press), pp. 561-574.
- Biesiada, E., Hamamori, Y., Kedes, L., and Sartorelli, V. (1999). Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac  $\alpha$ -actin promoter. *Mol Cell Biol* 19, 2577-84.
- Black, B. L., Molkentin, J. D., and Olson, E. N. (1998). Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interaction with MEF2. *Mol Cell Biol* 18, 69-77.



- Blackwell, T. K., Huang, J., Ma, A., Kretzner, L., Alt, F. W., Eisenman, R. N., and Weintraub, H. (1993). Binding of Myc proteins to canonical and non-canonical DNA sequences. *Mol. Cell Biol.* 13, 5216-5224.
- Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. *Science* 250, 1149-1151.
- Blackwell, T. K., and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* 250, 1104-1110.
- Bodis, S., Hemesath, T., and Fisher, D. E. (1997). Highly conserved asparagine in the basic domain of Myc is dispensable for DNA binding, transformation, and apoptosis. *Biochem Mol Med* 60, 102-7.
- Brennan, T. J., Chakraborty, T., and Olson, E. N. (1991). Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc Natl Acad Sci U S A* 88, 5675-9.
- Brown, E. J., and Schreiber, S. L. (1996). A signaling pathway to translational control. *Cell* 86, 517-520.
- Brownlie, P., Ceska, T., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* 5, 509-20.
- Carballo, E., Gilkeson, G. S., and Blackshear, P. J. (1997). Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (-/-) mice. Evidence that monocyte/macrophage progenitors may be responsible for TNFalpha overproduction. *J Clin Invest* 100, 986-995.
- Carballo, E., Lai, W. S., and Blackshear, P. J. (1998). Feedback inhibition of macrophage tumor necrosis factor- $\alpha$  production by tristetraprolin. *Science* 281, 1001-1005.
- Carey, J. (1991). Gel Retardation. *Methods in Enzymology* 208, 103-117.
- Chakraborty, T., Brennan, T. J., Li, L., Edmondson, D., and Olson, E. N. (1991). Inefficient homooligomerization contributes to the dependence of myogenin on E2A products for efficient DNA binding. *Mol Cell Biol* 11, 3633-41.
- Chandar, N., Billig, B., McMaster, J., and Novak, J. (1992). Inactivation of p53 gene in human and murine osteosarcoma cells. *Br. J. Cancer* 65, 208-214.
- Chen, C.-Y. A., and Shyu, A.-B. (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* 20, 465-470.
- Corps, A. N., and Brown, K. D. (1995). Insulin and insulin-like growth factor I stimulate expression of the primary response gene cMG1/TIS11b by a wortmannin-sensitive pathway in RIE-1 cells. *FEBS Lett* 368, 160-4.
- Cripps, R. M., Black, B. L., Zhao, B., Lien, C. L., Schulz, R. A., and Olson, E. N. (1998). The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during *Drosophila* myogenesis. *Genes Dev* 12, 422-34.

- Dang, C. V., Dolde, C., Gillison, M. L., and Kato, G. J. (1992). Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. USA* 89, 599-602.
- Davis, R. L., Cheng, P. F., Lassar, A. B., and Weintraub, H. (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* 60, 733-746.
- Davis, R. L., and Weintraub, H. (1992). Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* 256, 1027-30.
- DuBois, R. N., McLane, M. W., Ryder, K., Lau, L. F., and Nathans, D. (1990). A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *J. Biol. Chem.* 265, 19185-19191.
- Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S. C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8, 970-980.
- Evan, G. I., Wyllie, A. H., Gilbert, C. S., Littlewood, T. D., Land, H., Brooks, M., Waters, C. M., Penn, L. Z., and Hancock, D. C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell* 69, 119-128.
- Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363, 38-45.
- Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994). Structure and function of the b/HLH/Z domain of USF. *EMBO J* 13, 180-9.
- Fisher, D. E., Parent, L. A., and Sharp, P. A. (1993). High Affinity DNA-Binding Myc Analogs: Recognition by an  $\alpha$  Helix. *Cell* 72, 467-476.
- Fisher, F., Crouch, D. H., Jayaraman, P.-S., Clark, W., Gillespie, D. A. F., and Goding, C. R. (1993). Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs *in vivo*. *EMBO J* 12, 5075-5082.
- Gomperts, M., Corps, A. N., Pascall, J. C., and Brown, K. D. (1992). Mitogen-induced expression of the primary response gene cMG1 in a rat intestinal epithelial cell-line (RIE-1). *FEBS Lett* 306, 1-4.
- Gould, K. A., and Bresnick, E. H. (1998). Sequence determinants of DNA binding by the hematopoietic helix-loop-helix transcription factor TAL1: importance of sequences flanking the E-box core. *Gene Expr* 7, 87-101.
- Gross, A., McDonnell, J. M., and Korsmeyer, S. J. (1999). BCL-2 family members and the mitochondria in apoptosis. *Genes Dev* 13, 1899-911.
- Grumont, R. J., Rourke, I. J., and Gerondakis, S. (1999). Rel-dependent induction of A1 transcription is required to protect B cells from antigen receptor ligation-induced apoptosis. *Genes Dev.* 13, 400-411.
- Guedes, S., and Priess, J. (1997). The *C. elegans* MEX-1 protein is present in germline blastomeres and is a P granule component. *Development* 124, 731-739.

- Hafezi, F., Steinbach, J. P., Marti, A., Munz, K., Wang, Z. Q., Wagner, E. F., Aguzzi, A., and Reme, C. E. (1997). The absence of c-fos prevents light-induced apoptotic cell death of photoreceptors in retinal degeneration in vivo. *Nat. Med.* 3, 346-349.
- Halazonetis, T. D., and Kandil, A. N. (1992). Predicted structural similarities of the DNA binding domains of c-Myc and endonuclease Eco RI. *Science* 255, 464-466.
- Hamamori, Y., Wu, H. Y., Sartorelli, V., and Kedes, L. (1997). The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. *Mol Cell Biol* 17, 6563-73.
- Han, J., Wallen, H. D., Nunez, G., and White, E. (1998). E1B 19,000-molecular-weight protein interacts with and inhibits CED-4- dependent, FLICE-mediated apoptosis. *Mol. Cell. Biol.* 18, 6052-6062.
- Harfe, B. D., Branda, C. S., Krause, M., Stern, M. J., and Fire, A. (1998). MyoD and the specification of muscle and non-muscle fates during postembryonic development of the *C. elegans* mesoderm. *Development* 125, 2479-88.
- Harkin, D. P., Bean, J. M., Miklos, D., Song, Y. H., Truong, V. B., Englert, C., Christians, F. C., Ellisen, L. W., Maheswaran, S., Oliner, J. D., and Haber, D. A. (1999). Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* 97, 575-86.
- Hsieh, J.-K. X., Fredersdorf, S., Kouzarides, T., Martin, K., and Lu, X. (1997). E2F1-induced apoptosis requires DNA binding but not transactivation and is inhibited by the retinoblastoma protein through direct interaction. *Genes Dev.* 11, 1840-1852.
- Huang, J., Blackwell, T. K., Kedes, L., and Weintraub, H. (1996). Differences between MyoD DNA binding and activation site requirements revealed by a functional random sequence selection. *Mol. Cell Biol.* 16, 3893-3900.
- Huang, J., Weintraub, H., and Kedes, L. (1998). Intramolecular regulation of MyoD activation domain conformation and function. *Mol Cell Biol* 18, 5478-84.
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson Jr, J., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999). The transcriptional program in the response of human fibroblasts to serum. *Science* 283, 83-87.
- Janicke, R. U., Lee, F. H., and Porter, A. G. (1994). Nuclear c-Myc plays an important role in the cytotoxicity of tumor necrosis factor alpha in tumor cells. *Mol. Cell. Biol.* 14, 5661-5670.
- Jennings, B. H., Tyler, D. M., and Bray, S. J. (1999). Target specificities of *Drosophila* enhancer of split basic helix-loop-helix proteins. *Mol Cell Biol* 19, 4600-10.
- Juin, P., Hueber, A. O., Littlewood, T., and Evan, G. (1999). c-Myc-induced sensitization to apoptosis is mediated through cytochrome c release. *Genes Dev.* 13, 1367-81.
- Kanoh, J., Sugimoto, A., and Yamamoto, M. (1995). *Schizosaccharomyces pombe* *zfs1*<sup>+</sup> encoding a zinc-finger protein functions in the mating pheromone recognition pathway. *Mol. Biol. Cell* 6, 1185-1195.

- Kleefstrom, J., Vastrik, I., Saksela, E., Valle, J., Eilers, M., and Alitalo, K. (1994). c-Myc induces cellular susceptibility to the cytotoxic action of TNF-  $\alpha$ . *EMBO J.* 13, 5442-5450.
- Kowalik, T. F., DeGregori, J., Schwarz, J. K., and Nevins, J. R. (1995). E2F1 overexpression in quiescent fibroblasts leads to induction of cellular DNA synthesis and apoptosis. *J. Virol.* 69, 2491-2500.
- Lai, W. S., Carballo, E., Strum, J. R., Kennington, E. A., Phillips, R. S., and Blackshear, P. J. (1999). Evidence that Tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor  $\alpha$  mRNA. *Mol. Cell. Biol.* 19, 4311-4323.
- Lai, W. S., Stumpo, D. J., and Blackshear, P. J. (1990). Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. *J. Biol. Chem.* 265, 16556-16563.
- Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* 66, 305-15.
- Lee, J. E. (1997). Basic helix-loop-helix genes in neural development. *Curr Opin Neurobiol* 7, 13-20.
- Lemercier, C., To, R. Q., Carrasco, R. A., and Konieczny, S. F. (1998). The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of myoD. *EMBO J* 17, 1412-22.
- Ma, P. C., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77, 451-459.
- Ma, Q., and Herschman, H. R. (1995). The yeast homologue YTIS11, of the mammalian TIS11 gene family is a non-essential, glucose repressible gene. *Oncogene* 10, 487-494.
- Maleki, S. J., Royer, C. A., and Hurlburt, B. K. (1997). MyoD-E12 heterodimers and MyoD-MyoD homodimers are equally stable. *Biochemistry* 36, 6762-7.
- Meikrantz, W., and Schlegel, R. (1996). Suppression of apoptosis by dominant negative mutants of cyclin-dependent protein kinases. *J. Biol. Chem.* 271, 10205-10209.
- Mello, C. C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J. R. (1996). The PIE-1 protein and germline specification in *C. elegans* embryos. *Nature* 382, 710-712.
- Mesner, P. W., Epting, C. L., Hegarty, J. L., and Green, S. H. (1995). A timetable of events during programmed cell death induced by trophic factor withdrawal from neuronal PC12 cells. *J. Neurosci.* 15, 7357-7366.
- Michelson, A. (1996). A new turn (or two) for Twist. *Science* 272, 1449-1450.
- Molkentin, J. D., Black, B. L., Martin, J. F., and Olson, E. N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* 83, 1125-1136.
- Molkentin, J. D., and Olson, E. N. (1996). Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc Natl Acad Sci U S A* 93, 9366-73.

- Murray, M. V., Turnage, M. A., Williamson, K. J., Steinhauer, W. R., and Searles, L. L. (1997). The *Drosophila* suppressor of sable protein binds to RNA and associates with a subset of polytene chromosome bands. *Mol. Cell. Biol.* 17, 2291-2300.
- Murre, C., McCaw, P. S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56, 777-783.
- Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H., and Baltimore, D. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58, 537-544.
- Neuhold, L., and Wold, B. (1993). HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* 74, 1033-1042.
- Nevins, J. R. (1998). Toward an understanding of the functional complexity of the E2F and retinoblastoma families. *Cell Growth Differ* 9, 585-93.
- Olson, E. N., and Klein, W. H. (1994). bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev.* 8, 1-8.
- Pabo, C. O., and Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Ann. Rev. Biochem.* 61, 1053-1095.
- Parraga, A., Bellolell, L., Ferre-D'Amare, A. R., and Burley, S. K. (1998). Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* 6, 661-72.
- Phillips, A. C., Bates, S., Ryan, K. M., Helin, K., and Vousden, K. H. (1997). Induction of DNA synthesis and apoptosis are separable functions of E2F-1. *Genes Dev.* 11, 1853-1863.
- Phillips, A. C., Ernst, M. K., Bates, S., Rice, N. R., and Vousden, K. H. (1999). E2F-1 potentiates cell death by blocking antiapoptotic signaling pathways. *Mol. Cell* 4, 771-781.
- Postigo, A. A., and Dean, D. C. (1999). ZEB represses transcription through interaction with the corepressor CtBP. *Proc Natl Acad Sci U S A* 96, 6683-8.
- Ross, J. (1995). mRNA stability in mammalian cells. *Microbiol. Rev.* 59, 423-450.
- Rudner, D. Z., Breger, K. S., and Rio, D. C. (1998). Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *drosophila* subunit is dispensable in vivo. *Genes Dev.* 12, 1010-21.
- Schwarz, J. J., Chakraborty, T., Martin, J., Zhou, J. M., and Olson, E. N. (1992). The basic region of myogenin cooperates with two transcription activation domains to induce muscle-specific transcription. *Mol. Cell. Biol.* 12, 266-275.
- Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46, 659-667.
- Shi, Y., Glynn, J. M., Guilbert, L. J., Cotter, T. G., Bissonette, R. P., and Green, D. (1992). Role for c-Myc in activation-induced apoptotic cell death in T-cell hybridomas. *Science* 257, 212-214.

Shimizu, T. X., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997). Crystal structure of PHO4 bHLH domain-DNA complex: Flanking base recognition. *EMBO J* 16, 4689-4697.

Spolar, R. S., and Record Jr., M. T. (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263, 777-784.

Steitz, T. A. (1990). Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Quart. Rev. Biophys.* 23, 205-280.

Szymanski, P., and Levine, M. (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *EMBO Journal* 14, 2229-2238.

Tabara, H., Hill, R. J., Mello, C. C., Priess, J. R., and Kohara, Y. (1999). *pos-1* encodes a cytoplasmic zinc-finger protein essential for germline specification in *C. elegans*. *Development* 126, 1-11.

Taylor, G. A., Carballo, E., Lee, D. M., Lai, W. S., Thompson, M. J., Patel, D. D., Schenkman, D. I., Gilkeson, G. S., Broxmeyer, H. E., Haynes, B. F., and Blackshear, P. J. (1996). A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity* 4, 445-454.

Taylor, G. A., Thompson, M. J., Lai, W. S., and Blackshear, P. J. (1996). Mitogens stimulate the rapid nuclear to cytosolic translocation of tristetraprolin, a potential zinc-finger transcription factor. *Mol. Endocrinol.* 10, 140-146.

Taylor, G. A., Thompson, M. J., Lai, W. S., and Blackshear, P. L. (1995). Phosphorylation of tristetraprolin, a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase *in vitro*. *J. Biol. Chem.* 270, 13341-13347.

Thompson, M. J., Lai, W. S., Taylor, G. A., and Blackshear, P. J. (1996). Cloning and characterization of two yeast genes encoding members of the CCCH class of zinc finger proteins: zinc finger-mediated impairment of cell growth. *Gene* 174, 225-33.

Tronchere, H., Wang, J., and Fu, X.-D. (1997). A protein related to splicing factor U2AF<sup>35</sup> that interacts with U2AF<sup>65</sup> and SR proteins in splicing of pre-mRNA. *Nature* 388, 397-400.

Van Antwerp, M. E., Chen, D. G., Chang, C., and Prochownik, E. V. (1992). A point mutation in the MyoD basic domain imparts c-Myc-like properties. *Proc. Natl. Acad. Sci. USA* 89, 9010-9014.

Varnum, B. C., Lim, R. W., Sukhatme, V. P., and Herschman, H. R. (1989). Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 cells by the tumor promoter tetradecanoyl phorbol acetate. *Oncogene* 4, 119-120.

Varnum, B. C., Ma, Q., Chi, T., Fletcher, B., and Herschman, H. R. (1991). The TIS11 primary response gene is a member of a gene family that encodes proteins with a highly conserved sequence containing an unusual Cys-His repeat. *Mol. Cell. Biol.* 11, 1754-1758.

Voronova, A., and Baltimore, D. (1990). Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* 87, 4722-4726.

- Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V., and Baldwin, A. S., Jr. (1998). NF- $\kappa$ B antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. *Science* 281, 1680-1683.
- Warbrick, E., and Glover, D. (1994). A *Drosophila melanogaster* homolog of the *TIS11* family of immediate early genes that can rescue a *cdr1 cdc25* mutant strain of fission yeast. *Gene* 151, 243-246.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y., and Lassar, A. (1991). The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* 251, 761-766.
- Weintraub, H., Dwarki, V. J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A., and Tapscott, S. J. (1991). Muscle-specific transcriptional activation by MyoD. *Genes Dev.* 5, 1377-86.
- Weintraub, H., Genetta, T., and Kadesch, T. (1994). Tissue-specific gene activation by MyoD: determination of specificity by cis-acting repression elements. *Genes Dev* 8, 2203-11.
- Wendt, H., Thomas, R. M., and Ellenberger, T. (1998). DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem* 273, 5735-43.
- Worthington, M. T., Amann, B. T., Nathans, D., and Berg, J. M. (1996). Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proc. Natl. Acad. Sci. USA* 93, 13754-13759.
- Wu, M. X., Ao, Z., Prasad, K. V., Wu, R., and Schlossman, S. F. (1998). IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival. *Science* 281, 998-1001.
- Yin, Z., Xu, X. L., and Frasch, M. (1997). Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* 124, 4971-82.
- Zhan, Y., Cleveland, J. L., and Stevens, J. L. (1997). A role for c-myc in chemically induced renal-cell death. *Mol. Cell. Biol.* 17, 6755-6764.
- Zong, W. X., Edelstein, L. C., Chen, C., Bash, J., and Gelinas, C. (1999). The prosurvival Bcl-2 homolog Bfl-1/A1 is a direct transcriptional target of NF-kappaB that blocks TNFalpha-induced apoptosis. *Genes Dev.* 13, 382-387.

Figure 1

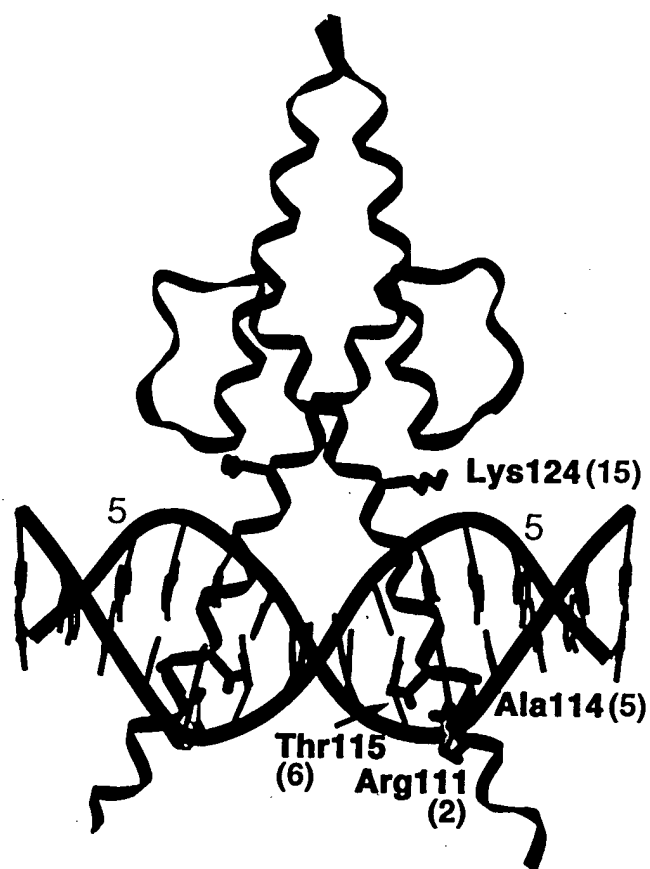




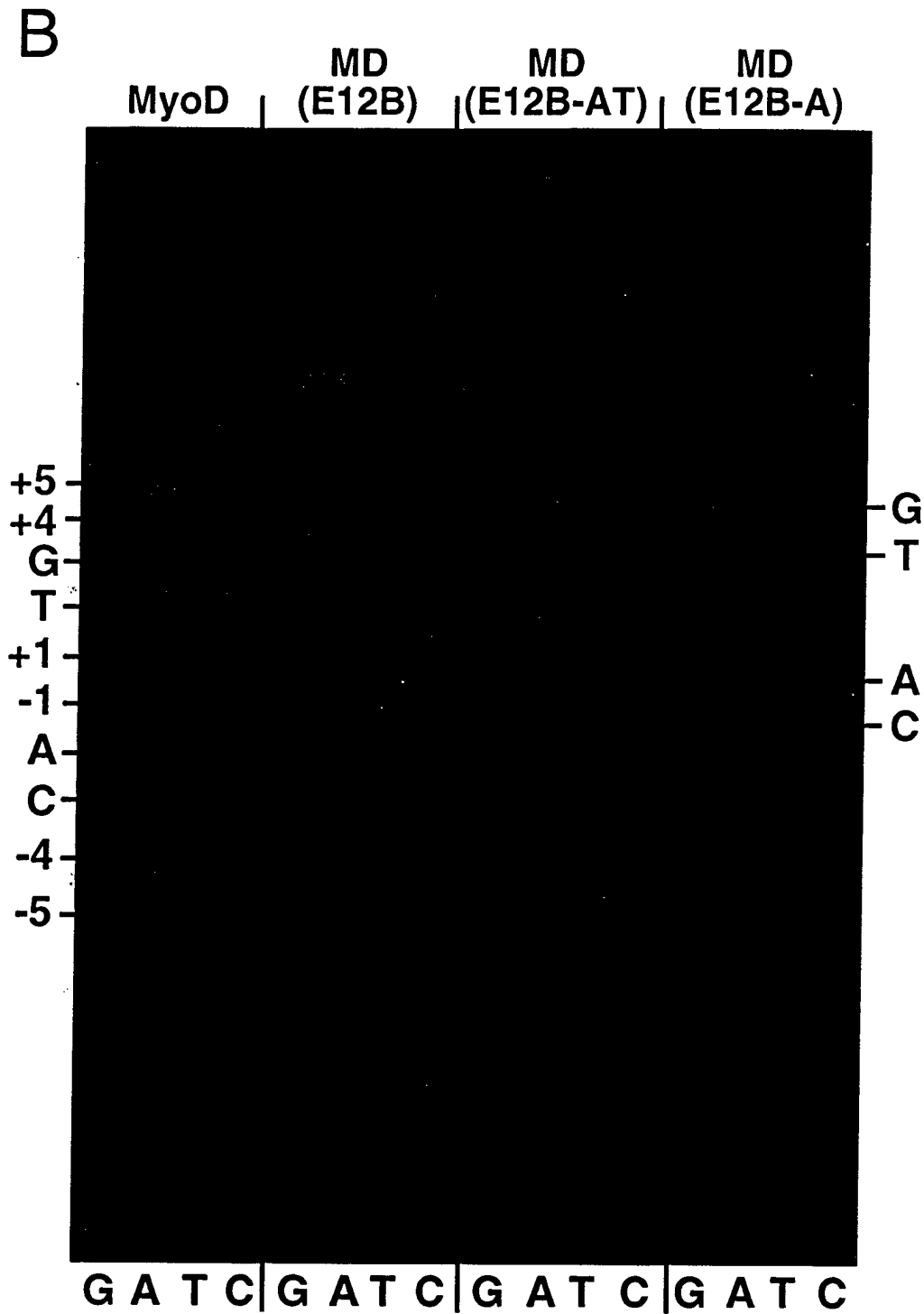
Figure 2

<u>BASIC REGIONS:</u>		<u>MUSCLE:</u>	
MyoD	K R K T T N A D R R K A A T M R E R R R L S K V	++++	
E12	Q K A E R E K E R R V A N N A R E R L R V R D I	No	
Twist	Q S Y E E L Q T R V M A N V R E R Q R T Q S L	ND	
<u>MUTANTS:</u>			
MD (E12B)	[Q K A E R E K E R R V A N N A R E R L R L S K V]	No	
MD (E12B-A)	[Q K A E R E K E R R V A A N A R E R L R L S K V]	No	
MD (E12B-AT)	[Q K A E R E K E R R V A A T A R E R L R L S K V]	+	
MD (E12BJ)	[Q K A E R E K E R R V A N N A R E R L R V R D I]	No	
E12 (MDB)	Q K A [T T N A D R R K A A T M R E R R R V R D I]	No	
E12 (MDBJ)	Q K A [T T N A D R R K A A T M R E R R R L S K V]	++	
E12 (AT, MDJ)	Q K A E R E K E R R V A A T A R E R L R L S K V	+	
E12 (AT, K)	Q K A E R E K E R R V A A T A R E R L R V R K I	+	
E12 (AT)	Q K A E R E K E R R V A A T A R E R L R V R D I	No	
	1 5 10 15		
	Basic		Junction

Figure 3A

A

● ● ● ●  
 -5 -4 -3 -2 -1 1 2 3 4 5  
 5' B-ACCCCCNNCANNTGNNGCCTGAT-A 3' D3  
 5' B-ACCCCCNNNNCANNTGNNNNCTGAT-A 3' D6



C

<u>Protein</u>	<u>Consensus</u>	<u>Center</u>	<u>Flank</u>
MD	$\begin{array}{c} \text{G} \\ \text{A} \end{array} \text{ACAGCTGT} \begin{array}{c} \text{T} \\ \text{C} \end{array}$	MD	MD/MD
E2A	$\text{NT} \begin{array}{c} \text{CA} \\ \text{GG} \end{array} \text{CC} \text{TGAN}$	E2A	E2A/E2A
MD(E12B)	$\begin{array}{c} \text{G} \\ \text{A} \end{array} \text{CCATATGG} \begin{array}{c} \text{T} \\ \text{C} \end{array}$	Tw1	Tw1/Tw1
MD(E12B-A)	$\begin{array}{c} \text{GC} \\ \text{AA} \end{array} \text{CATATG} \begin{array}{c} \text{TT} \\ \text{GC} \end{array}$	Tw1	Tw1/Tw1
MD(E12B-AT)	$\begin{array}{c} \text{G} \\ \text{A} \end{array} \text{ACAGCTGT} \begin{array}{c} \text{T} \\ \text{C} \end{array}$ -5 +5	MD	MD/MD

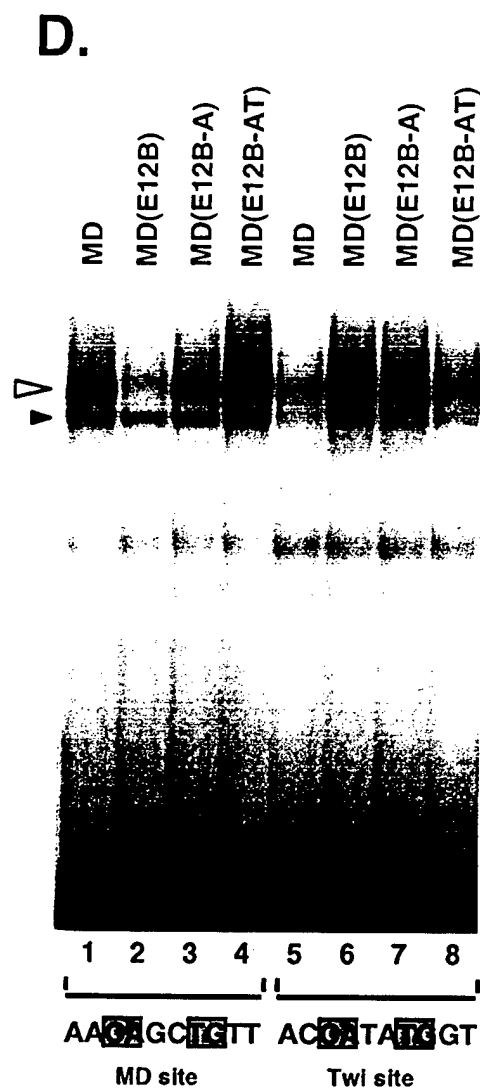
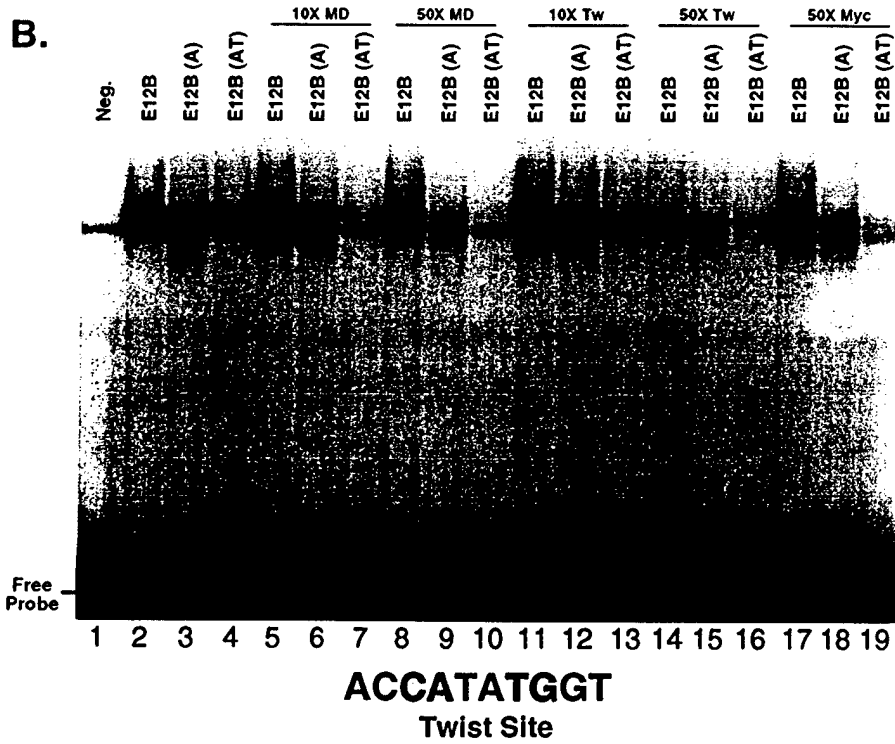
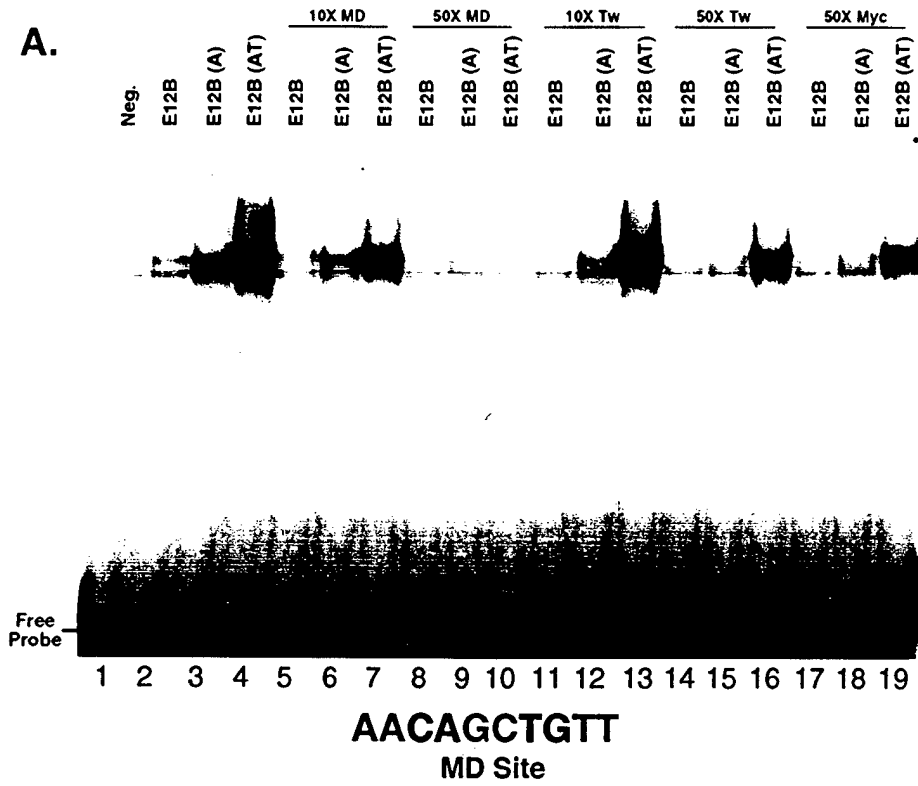


Figure 4



A

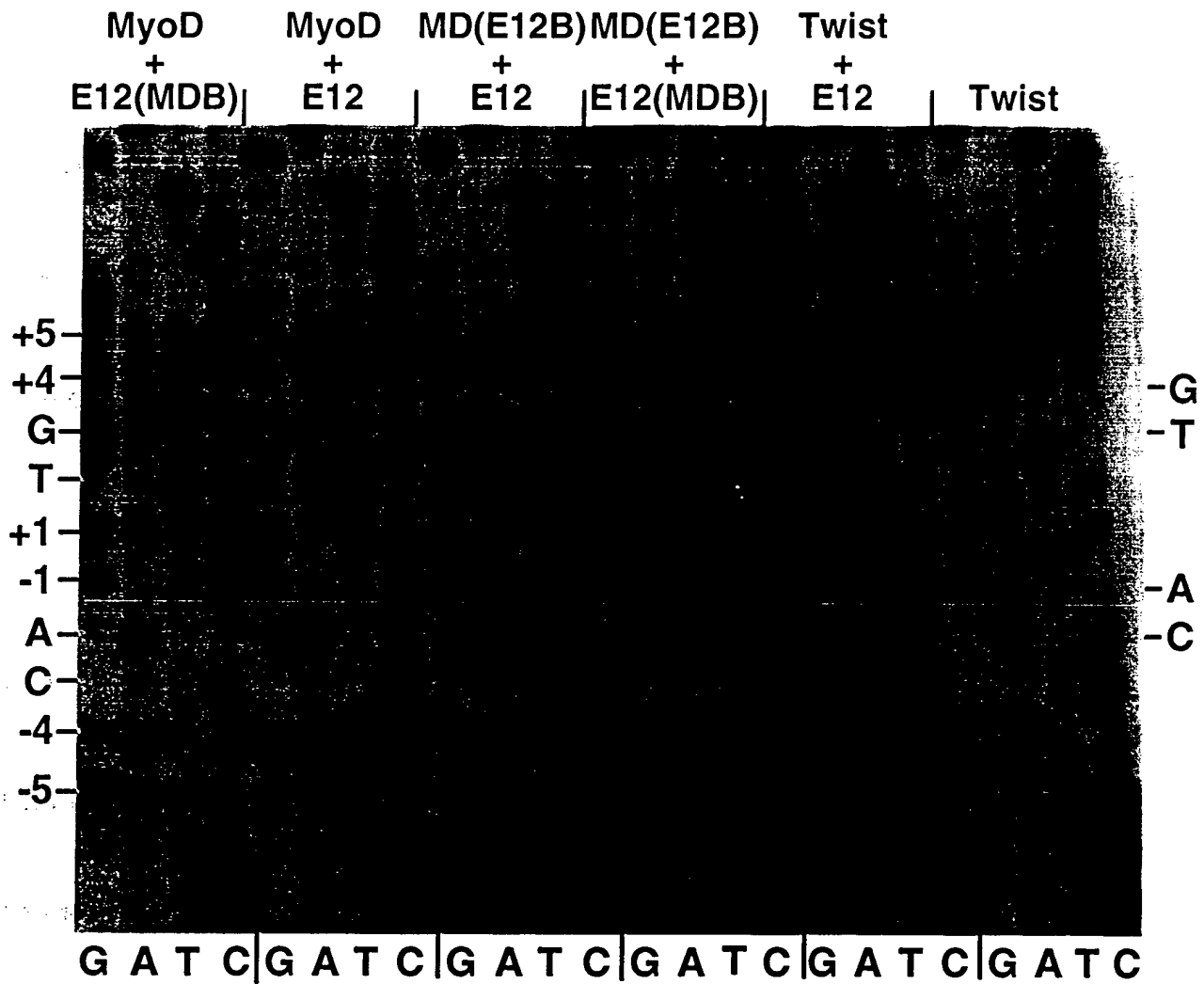


Figure 5B

## B

<u>Protein</u>	<u>Consensus</u>	<u>Center</u>	<u>Flank</u>
MD+E12	NTCA <sup>CC</sup> TTGT <sup>T</sup> GG      C	E2A	E2A/MD
MD (E12B) +E12	NTCA <sup>CC</sup> TTGAN GG      C	E2A	E2A/E2A
MD+E12 (MDB)	GACA <sup>CC</sup> TTGT <sup>T</sup> GG      C	E2A	MD/MD
MD (E12B) +E12 (MDB)	GC <sup>CA</sup> TA <sup>TG</sup> TT AA      GG      C	Tw1	Tw1/Tw1
Tw1+E12	GC <sup>CA</sup> TA <sup>TG</sup> TT AA      GG      C -5                 +5	Tw1	Tw1/Tw1



C

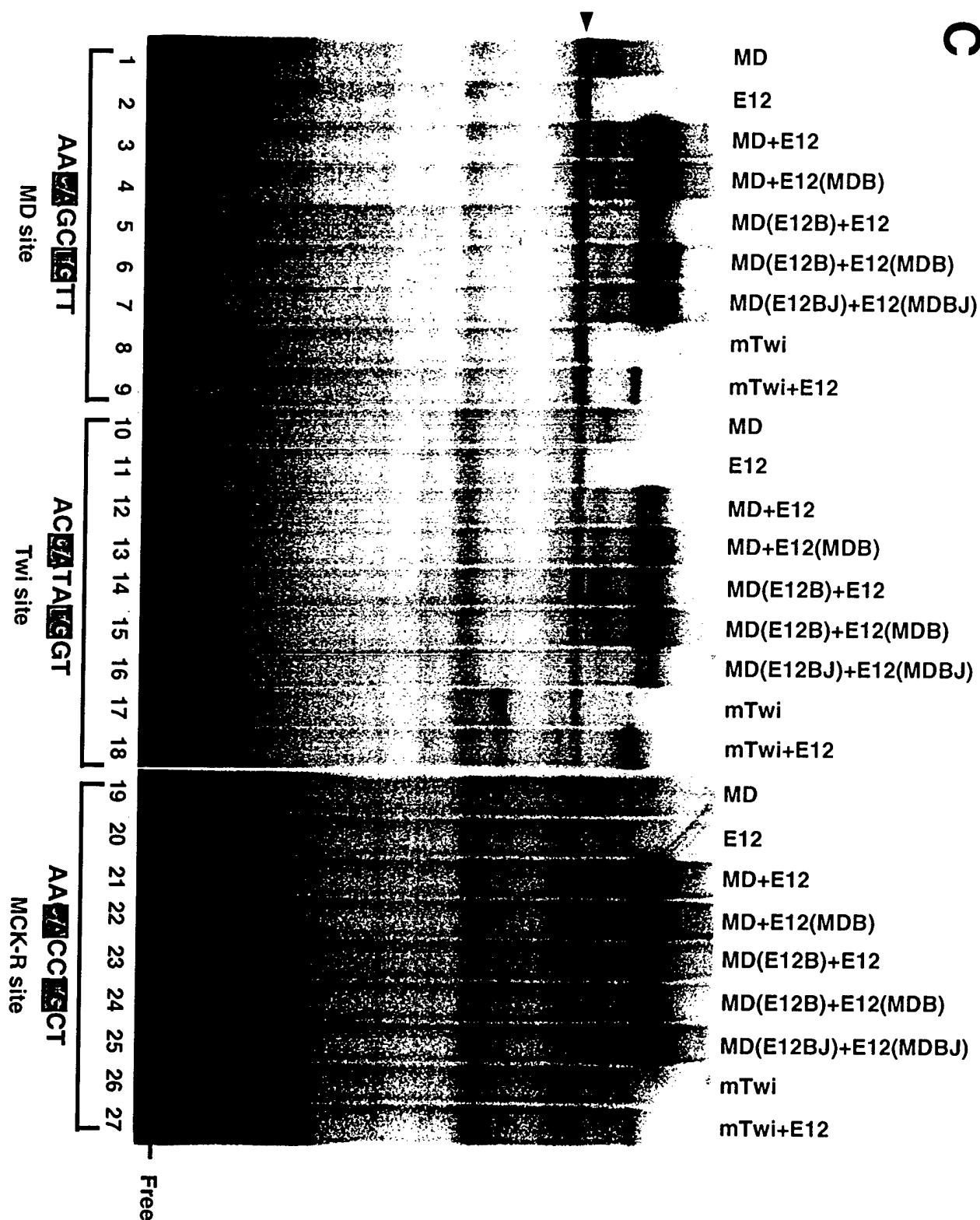


Figure 5C

Figure 6

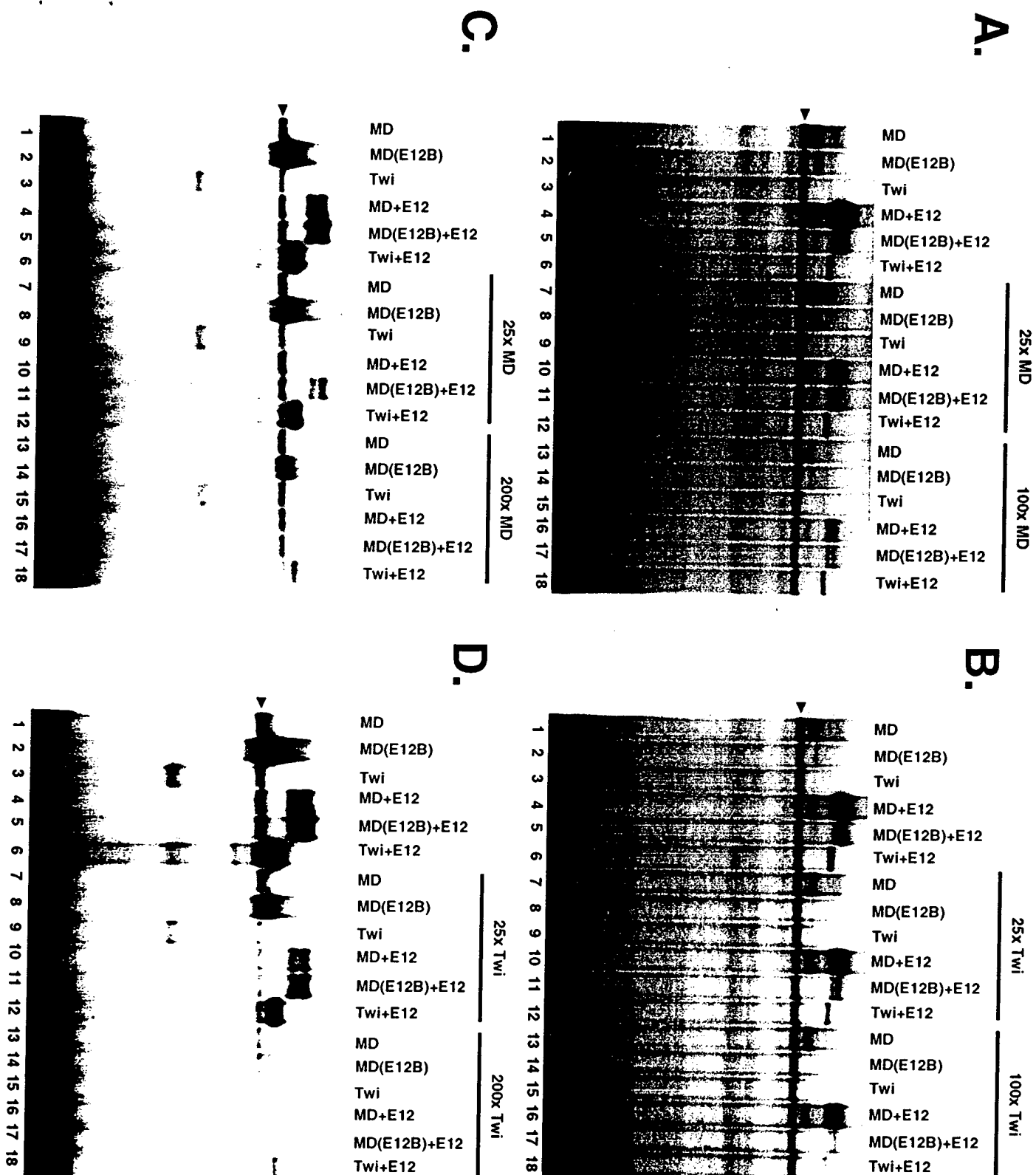


Figure 7

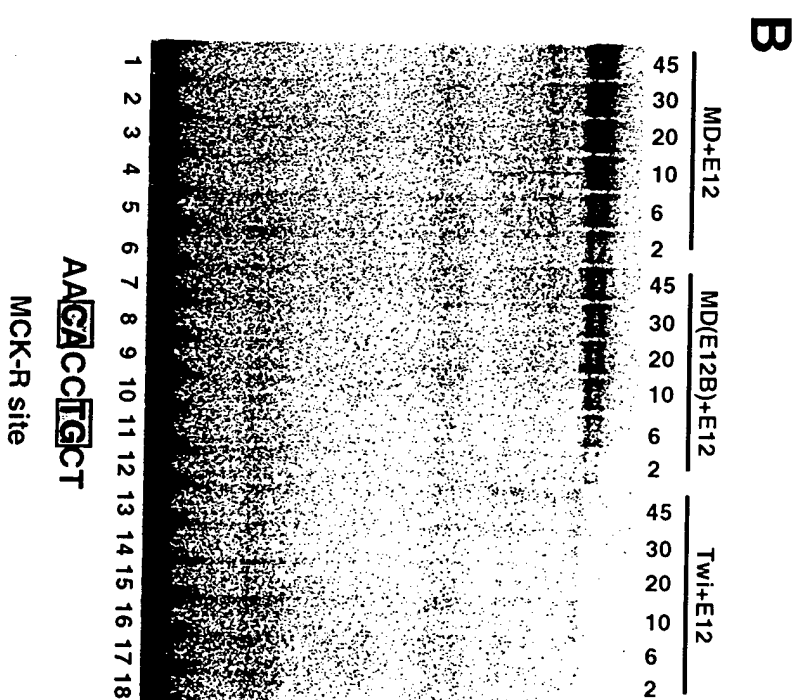
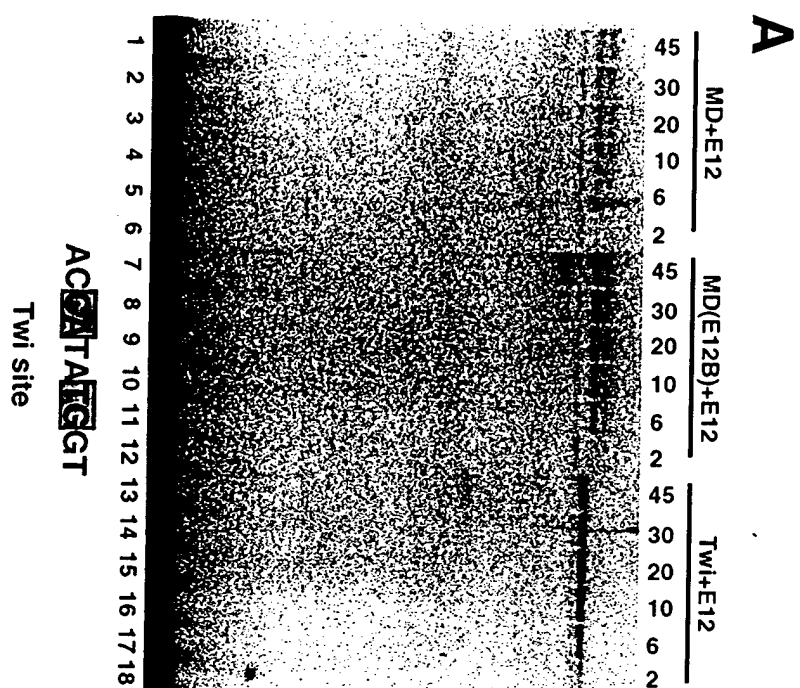


Figure 8A

A

	1	5	10	15
MyoD	R R K A A T M R E R R R L S K V			
E12	R R V A N N A R E R L R V R D I			
Twist	Q R V M A N V R E R Q R T Q S L			
MD-AAATA	R R <u>A A A T A</u> R E R R R L S K V			
MD-AAANA	R R <u>A A A N A</u> R E R R R L S K V			
MD-AANNA	R R <u>A A N N A</u> R E R R R L S K V			
MD-(AK)	R R K A A T M R E R R R L <u>A</u> K V			
MD-(AA)	R R K A A T M R E R R R L <u>A A</u> V			
MD-(AD)	R R K A A T M R E R R R L <u>A D</u> V			
MD-(AS)	R R K A A T M R E R R R L <u>A S</u> V			
MD-(QS)	R R K A A T M R E R R R L <u>Q S</u> V			
MD-AAATA(AK)	R R <u>A A A T A</u> R E R R R L <u>A</u> K V			
MD-AAATA(AA)	R R <u>A A A T A</u> R E R R R L <u>A A</u> V			
MD-AAANA(AK)	R R <u>A A A N A</u> R E R R R L <u>A</u> K V			
MD-AAANA(AA)	R R <u>A A A N A</u> R E R R R L <u>A A</u> V			
MD-AAANA(AD)	R R <u>A A A N A</u> R E R R R L <u>A D</u> V			
MD-AAANA(QS)	R R <u>A A A N A</u> R E R R R L <u>Q S</u> V			
MD-AANNA(AD)	R R <u>A A N N A</u> R E R R R L <u>A D</u> V			

Basic Region

Junction

Figure 8B

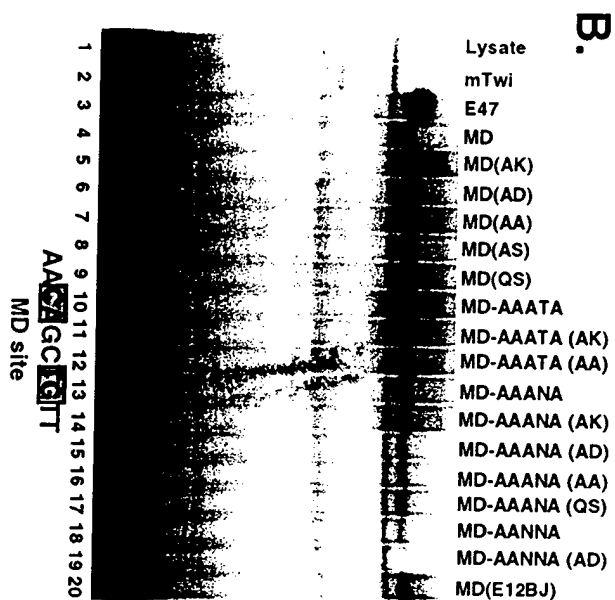


Figure 8C

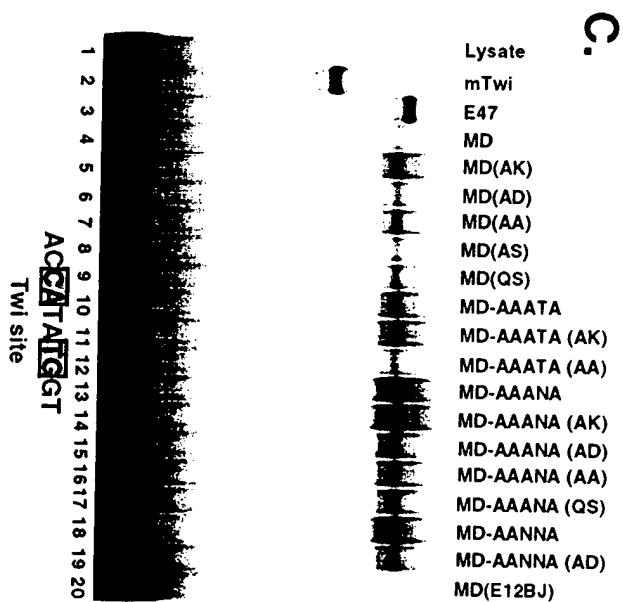


Figure 9

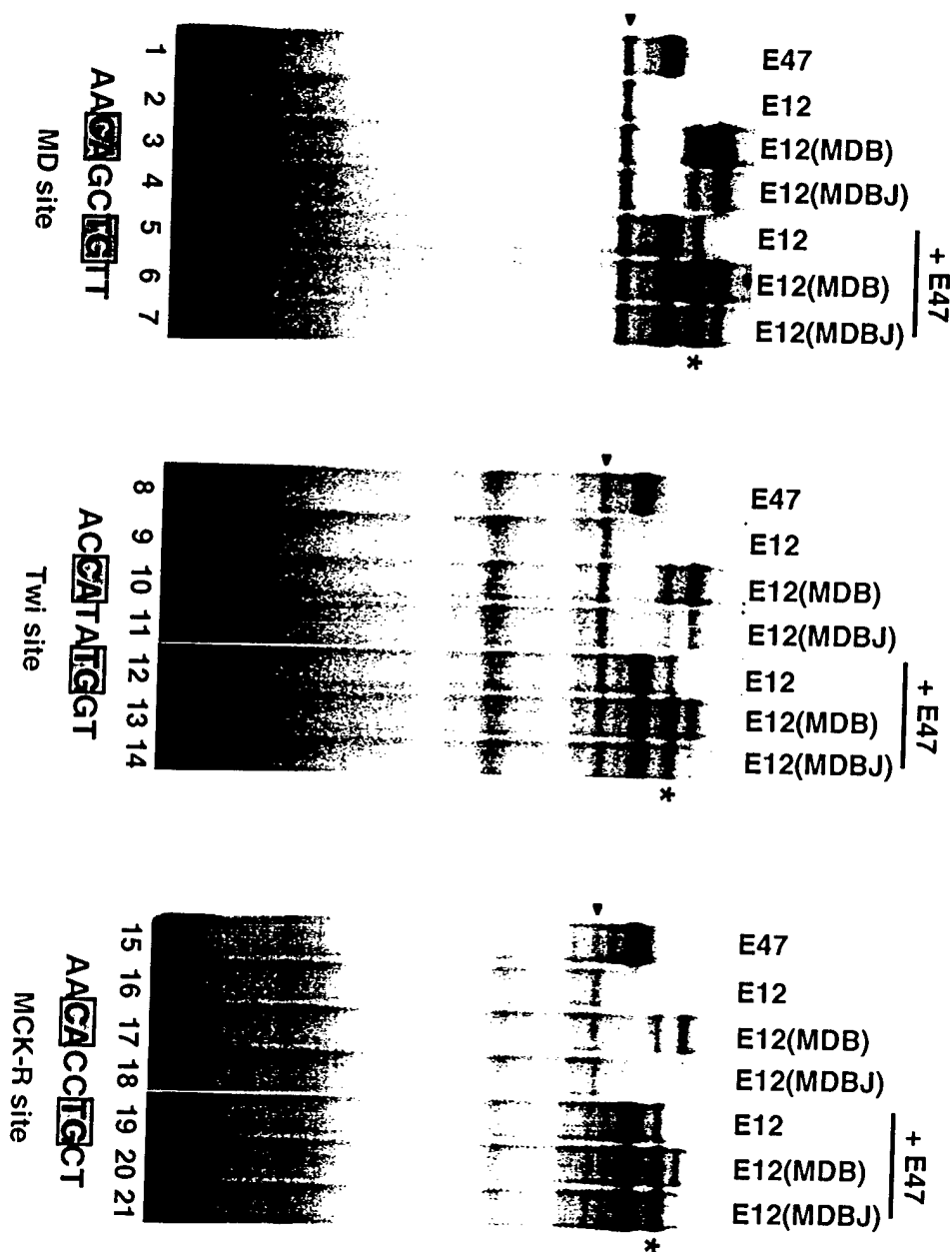
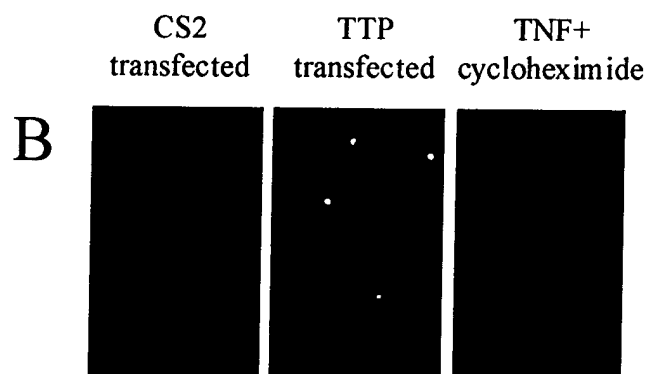
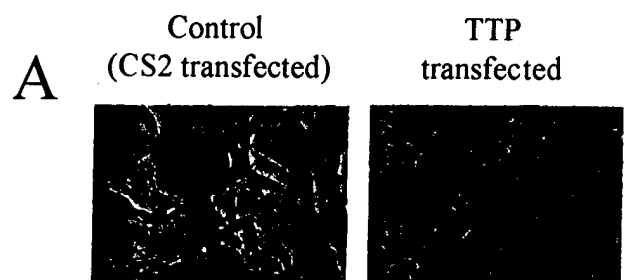


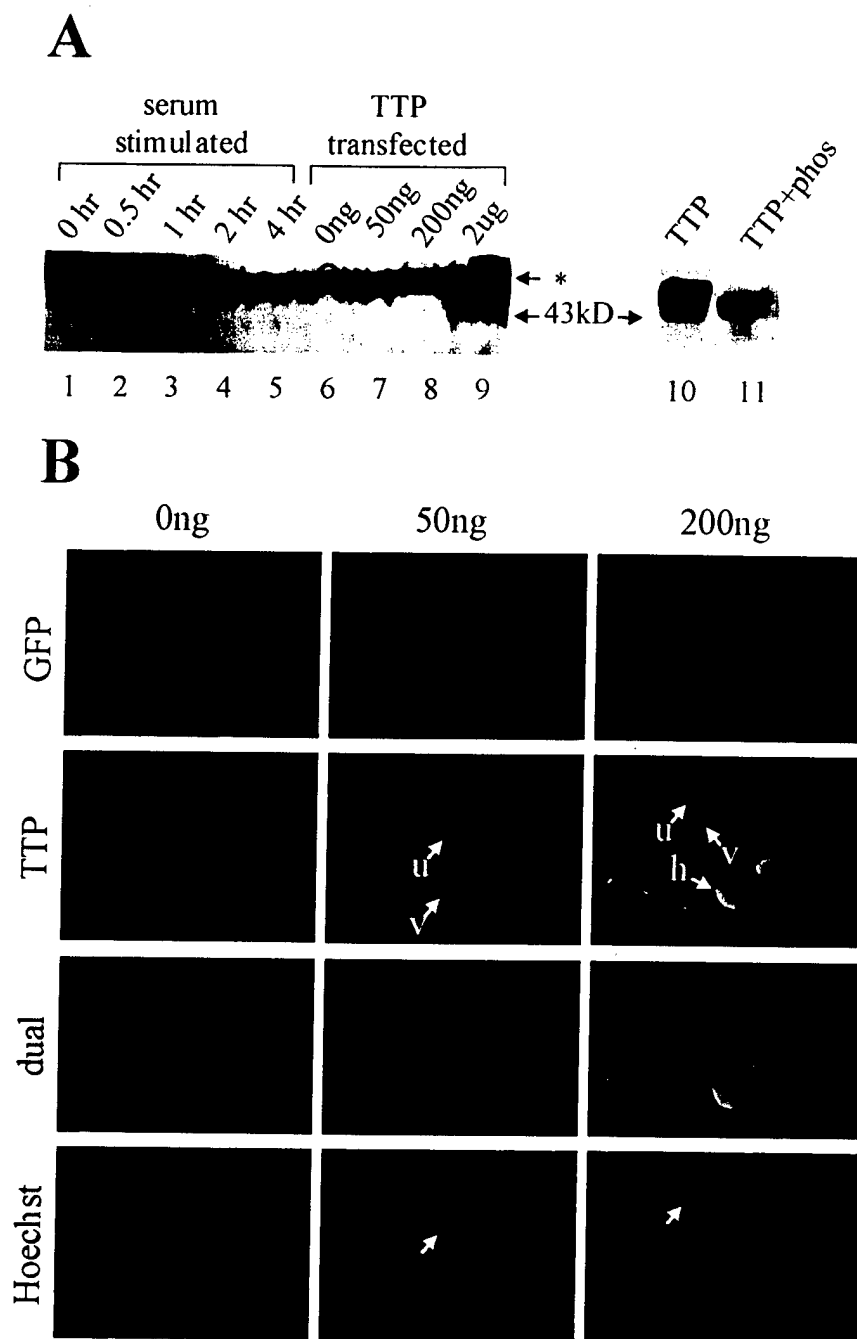
Figure 10



**C**

cell line	percentage apoptotic blue cells			
	24hr		48hr	
	CS2	TTP	CS2	TTP
3T3	1 ± 1	49 ± 7	10 ± 1	95 ± 2
HeLa	3 ± 2	39 ± 4	12 ± 5	78 ± 9
U2OS	2 ± 1	15 ± 4	13 ± 5	73 ± 5
SAOS2	0 ± 0	3 ± 4	10 ± 3	57 ± 15
293	3 ± 1	10 ± 3	7 ± 2	40 ± 4
HFK*	11 ± 2	38 ± 8	14 ± 4	70 ± 8
MEF*	17 ± 3	56 ± 7	18 ± 2	88 ± 4

Figure 11



**C**

			TTP staining level		
TTP	morphology	Tot.	Undetect.	Vis.	High
50ng	apoptotic	21%	72%	18%	10%
	non-apoptotic	79%	71%	15%	14%
200ng	apoptotic	58%	38%	41%	21%
	non-apoptotic	42%	51%	25%	24%



Figure 12

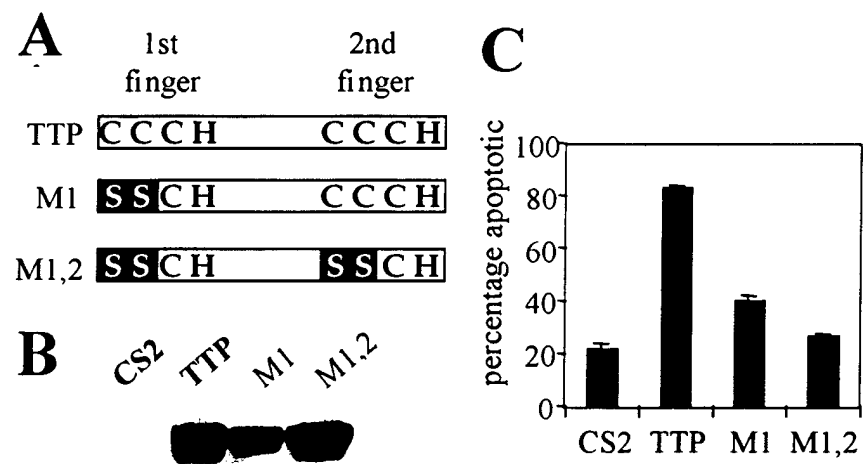
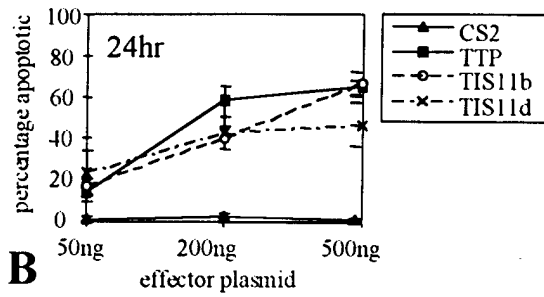
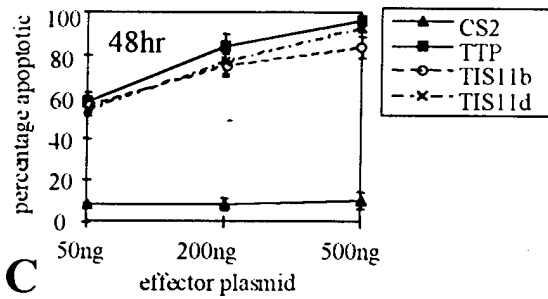


Figure 13

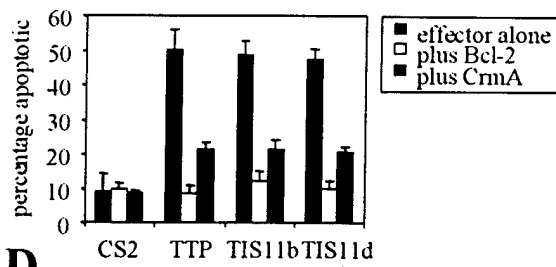
**A**



**B**



**C**



**D**

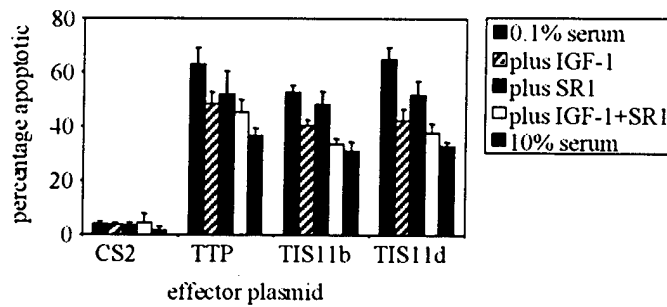
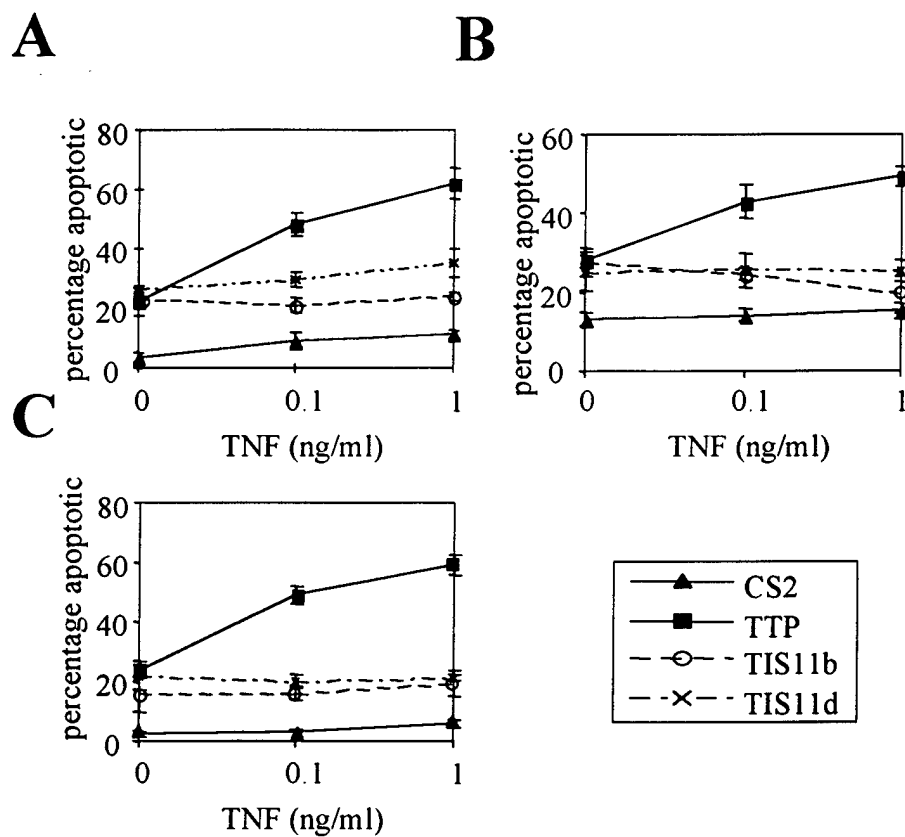


Figure 14



## FINAL REPORT

### Manuscripts:

Huang J, **Blackwell TK**, Kedes L, Weintraub H. Differences between MyoD DNA binding and activation site requirements revealed by a functional random sequence selection. *Mol Cell Biol* 1996;16:3893-00.

**Blackwell TK**, Kophengnavong T, Johnson BA. Biochemical and functional analyses of c-Myc- and TTP-related proteins. Abstract for 1997 Era of Hope meeting.

De J, Lai WS, Thorn JM, Goldsworthy SM, Liu X, **Blackwell TK**, Blackshear PJ. Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression. *Gene* 1999;228:133-45.

Kophengnavong T, Michnowicz JE, **Blackwell TK**. Establishment of Distinct MyoD, E2A, and Twist DNA binding specificities by different basic region-DNA conformations. *Mol Cell Biol* 2000; 20: 261-272.

Johnson B, Geha M, **Blackwell TK**. Similar but Distinct Effects of the Tristetraprolin/TIS11 Immediate-Early Proteins on Cell Survival. *Oncogene* 2000, pending minor revisions.

### Personnel:

Blackwell, T. Keith  
Michnowicz, Jennifer E.  
Johnson, Barbra A.  
Liu, Xiaoying  
Harford, Catherine  
Kophengnavong, Thiphaphone

## Differences between MyoD DNA Binding and Activation Site Requirements Revealed by Functional Random Sequence Selection

JING HUANG,<sup>1,2</sup> T. KEITH BLACKWELL,<sup>3</sup> LARRY KEDES,<sup>2\*</sup> AND HAROLD WEINTRAUB<sup>1</sup>

*Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104<sup>1</sup>; Institute for Genetic Medicine and Department of Biochemistry and Molecular Biology, USC School of Medicine, Los Angeles, California 90033<sup>2</sup>; and Center for Blood Research and Harvard Medical School, Boston, Massachusetts 02115<sup>3</sup>*

Received 14 December 1995/Returned for modification 6 February 1996/Accepted 26 March 1996

**A method has been developed for selecting functional enhancer/promoter sites from random DNA sequences in higher eukaryotic cells. Of sequences that were thus selected for transcriptional activation by the muscle-specific basic helix-loop-helix protein MyoD, only a subset are similar to the preferred *in vitro* binding consensus, and in the same promoter context an optimal *in vitro* binding site was inactive. Other sequences with full transcriptional activity instead exhibit sequence preferences that, remarkably, are generally either identical or very similar to those found in naturally occurring muscle-specific promoters. This first systematic examination of the relation between DNA binding and transcriptional activation by basic helix-loop-helix proteins indicates that binding *per se* is necessary but not sufficient for transcriptional activation by MyoD and implies a requirement for other DNA sequence-dependent interactions or conformations at its binding site.**

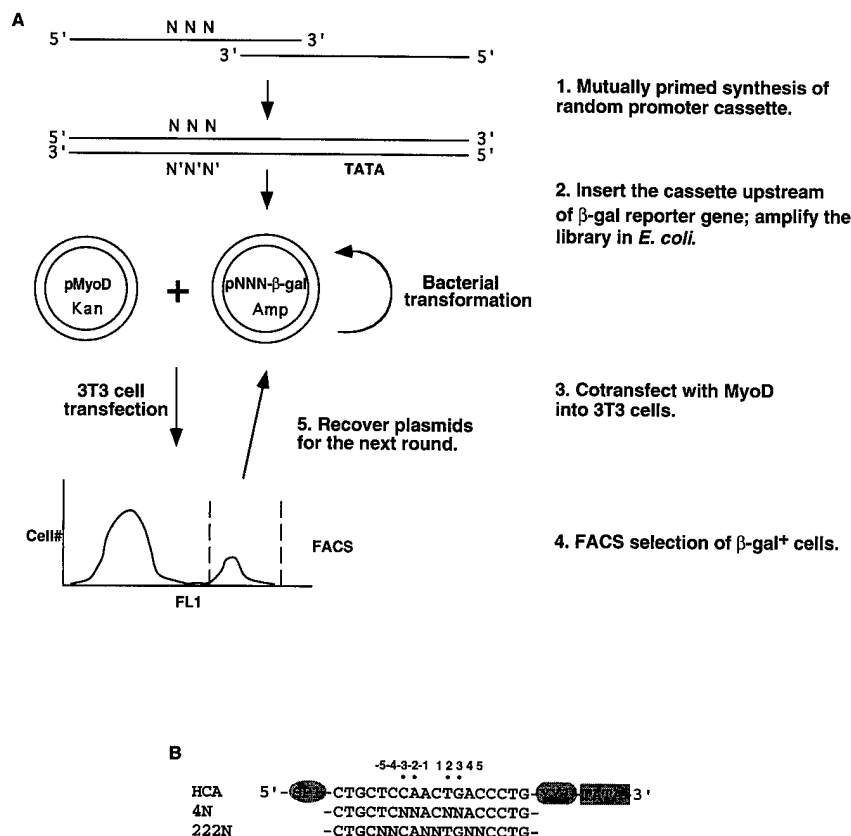
The basic helix-loop-helix (bHLH) family of proteins (48) includes many that are pivotal in cell type determination, such as the MyoD group (including MyoD, myogenin, myf-5, and MRF4) for skeletal myogenesis (10, 20, 55, 73, 75). The bHLH proteins form homo- or heterodimers through the HLH domain and bind to DNA through the adjacent basic region (BR) (16, 72). They bind as dimers to DNA sites that generally share the consensus CANNTG (the E box) (8, 49), with each BR occupying half of the site (8, 19, 21, 44).

MyoD family members form homodimers relatively inefficiently (67) and instead appear to activate transcription of muscle-specific genes as heterodimers with the widely expressed E proteins (including E2A [E12 and E47], E2-2, and HEB [3]) (41). E proteins also form heterodimers with other tissue-specific bHLH proteins (49), to regulate still different sets of genes in various tissues such as erythrocytes (1), pancreas (51), and neurons (36), and they form homodimers which are essential for B-cell differentiation (4, 18, 39). The ability of these different protein complexes to regulate different sets of genes appears paradoxical, because their DNA binding specificities are often surprisingly overlapping. For example, certain sites that can be bound well by either MyoD-E or E-E complexes *in vitro* (8, 49) can be activated by only one or the other of these complexes in the cell (25, 77). It therefore appears either that other protein factors are involved in determining transcriptional specificity or that subtle differences in DNA binding specificity of the bHLH proteins must be biologically significant. These two mechanisms may not be mutually exclusive. Evidence suggests that MyoD may require a positive cofactor to activate transcription of appropriate genes (9, 16, 17, 47, 76) and that it can be prevented from acting at inappropriate sites by repressors that recognize bases overlapping the E box (25, 77). Remarkably, these proposed cofactor and repressor functions both appear to depend on the presence of

particular MyoD BR residues, mutations of which may allow DNA binding but not transcriptional activity. X-ray crystallographic studies indicate that in the bound complex these MyoD residues are in intimate proximity to the DNA and that the MyoD BR assumes a binding conformation different from that of the E47 protein (19, 44), suggesting that this conformation may in turn constitute an important aspect of interactions between MyoD (or other bHLH proteins) and coactivator or repressor functions at the promoter.

In light of these observations, it is critical to study systematically the DNA sequence requirements for transcriptional activation by bHLH proteins. Preferred binding sites for MyoD and E proteins have been identified by random-sequence selection *in vitro* (8, 67, 79), but such assays cannot predict the ability of these proteins to support transcriptional activation. To this end, we have developed the first system that allows proteins to select functional DNA targets from random sequences in mammalian cells. In this new method, dubbed the selection of *in vivo* target elements (SITE) technique (Fig. 1A), oligonucleotides with random sequences at positions of interest were embedded in a muscle-specific promoter cassette, in this case the human cardiac  $\alpha$ -actin (HCA) enhancer/promoter cassette, located upstream of a  $\beta$ -galactosidase ( $\beta$ -Gal) reporter gene. The resulting plasmid library was cotransfected with a MyoD expression vector into murine NIH 3T3 fibroblasts, cells that expressed  $\beta$ -Gal were then selected by fluorescence-activated cell sorting (FACS), and finally plasmid DNA was extracted from selected cells and amplified in *Escherichia coli* to allow this selection procedure to be repeated. Three such selection rounds yielded sequences through which MyoD can activate transcription *in vivo*. These sequences overlap with but are generally distinct from optimal MyoD or MyoD-E binding sites identified *in vitro* (8, 79). Instead, the selected sites are similar to various functional E boxes present in native muscle-specific promoters. Remarkably, the active sequences selected from the HCA promoter context do not necessarily activate transcription in the context of another muscle-specific promoter, such as the muscle creatine kinase (MCK) promoter. These findings indicate, surprisingly, that

\* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, USC School of Medicine, HMR 413, 2011 Zonal Ave., Los Angeles, CA 90033. Phone: (213) 342-1145. Fax: (213) 342-2764.



**FIG. 1.** Selection scheme. (A) An oligonucleotide was synthesized with random sequences (designated N) at positions of interest in the protein-binding site. A double-stranded cassette was generated by extension over a complementary segment and cut with appropriate restriction enzymes at both ends. The cassette was then ligated unidirectionally into a  $\beta$ -Gal reporter plasmid such that transcription of the  $\beta$ -Gal gene is under the control of the randomized enhancer/promoter cassette. The ligation mixture was electroporated into *E. coli* cells, the transformants were plated on LB-agarose-ampicillin plates and allowed to form colonies, which were then pooled, and the plasmids were extracted. (The size of the plasmid library was made larger than  $4.6 \times 4^n$ ,  $n$  being the number of nucleotides made random, to ensure a probability of no less than 99% to encompass the complexity of the library [59].) The plasmid pool was cotransfected with a MyoD expression plasmid into 3T3 cells, and  $\beta$ -Gal<sup>+</sup> cells were collected by using a fluorescence-activated cell sorter, the stringency of selection being controlled by fluorescence gating. The plasmids in the collected cells were retrieved by the Hirt protocol, amplified by transformation back into *E. coli* cells, and used for the next round of selection. After the desired number of rounds of selection, the active enhancer sequences were determined by DNA sequencing and confirmed by a transfection test. (B) Core sequences of the random cassettes 4N and 222N. The 18-bp core sequences of 4N and 222N are based on the MyoD binding site of the HCA promoter (46). The consensus CA-TG motif is indicated by dots.

the optimal binding sites are not always capable of activating transcription and suggest that binding per se is not sufficient for activation, which may thus be very sensitive to promoter context and/or binding conformation.

#### MATERIALS AND METHODS

**Plasmid constructions.** The reporter plasmid pool pNNN- $\beta$ -gal (Fig. 1A) was constructed by ligating HCA promoter region -89 to +68 (46) containing each random nucleotide cassette (Fig. 1B), the *lacZ* reporter gene from the pNASS $\beta$  vector (Clontech), and a fragment from the pBluescript II KS (+) vector (Stratagene) that contained the ampicillin resistance gene and the ColE1 bacterial replication origin. Two E boxes (*PvuII* sites) in the plasmid backbone were deleted to allow unambiguous analysis. The MyoD expression vector used for SITE experiments, pMyoD-Kan (Fig. 1A), was derived from the pEMSV-MyoD expression vector (16) by inserting the kanamycin resistance gene at the *ScaI* site of the ampicillin resistance gene, which inactivates the latter selection marker. In plasmid 115MCK- $\beta$ -gal, the MCK enhancer region (-1207 to -1093) and promoter region (-82 to +7) were used to drive the *lacZ* reporter gene in the same plasmid backbone as in pNNN- $\beta$ -gal. The various E-box substitutions into the MCK context (Fig. 4) were created by mutagenesis using PCR (31) and confirmed by DNA sequence analysis.

**Cell culture and transfection.** NIH 3T3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% bovine calf serum (Gibco). Cells were cotransfected with 7.5  $\mu$ g of  $\beta$ -Gal reporter plasmid and 7.5  $\mu$ g of MyoD expression vector per 10-cm-diameter plate by calcium phosphate coprecipitation (12). Seventeen to 21 h posttransfection, the cells were switched to differentiation medium (Dulbecco's modified Eagle's medium containing 2%

horse serum [Gibco] plus transferrin [Sigma; 10  $\mu$ g/ml] and insulin [Sigma; 10  $\mu$ g/ml]) for 2 days before they were harvested for FACS.

**FACS selection of  $\beta$ -Gal<sup>+</sup> cells.** Cells were washed three times with 1 $\times$  phosphate-buffered saline (PBS), detached and dissociated in 5 mM EDTA in PBS, and passed through a 23-gauge needle to further dissociate doublets or bigger clumps. (EDTA was used instead of trypsin because plasma membrane damage caused by trypsin would increase leakage during and after the hypotonic shock.) The cells were resuspended in 50  $\mu$ l of 1 $\times$  PBS at  $1 \times 10^7$  to  $5 \times 10^7$  cells per ml, transferred to Falcon 2058 tubes, and warmed to 37°C. Flow cytometry analysis of  $\beta$ -Gal activity of the cells was performed as described previously (22, 53). Briefly, 50  $\mu$ l of 2 mM fluorescein di- $\beta$ -D-galactoside (FDG; Molecular Probes) in H<sub>2</sub>O was added to the cells, and the mixture was incubated at 37°C for 1 min. At the end of the incubation, 10 volumes (1 ml) of ice-cold 1 $\times$  PBS was immediately added, and the reaction mixture was placed on ice for 20 to 60 min. The reaction was terminated by addition of 2 mM phenylethyl- $\beta$ -D-thiogalactoside (Molecular Probes). Prior to sorting, propidium iodide (Sigma) was added (5  $\mu$ g/ml). Cells were analyzed and sorted at 4°C on a FACS machine (Vantage; Becton Dickinson). Fluorescein isothiocyanate (FITC)-positive cells (in which FDG had been converted to fluorescein) were collected into a 35-mm-diameter Falcon dish coated with 1% SeaPlaque agarose in 1 $\times$  PBS.

**Recovery of plasmids from FACS-selected cells.** Plasmids in the sorted cells were extracted by the Hirt protocol (32). Electroporation of highly competent *E. coli* cells was used for efficient recovery of selected plasmids (29). Since the *lacZ* reporter plasmid and the MyoD expression vector were constructed to carry different bacterial selection markers (Fig. 1A), the reporter plasmids can be selectively amplified in *E. coli*. The bacterial transformants were pooled, and plasmids were extracted from them (Maxiprep; Qiagen). In the final round of the

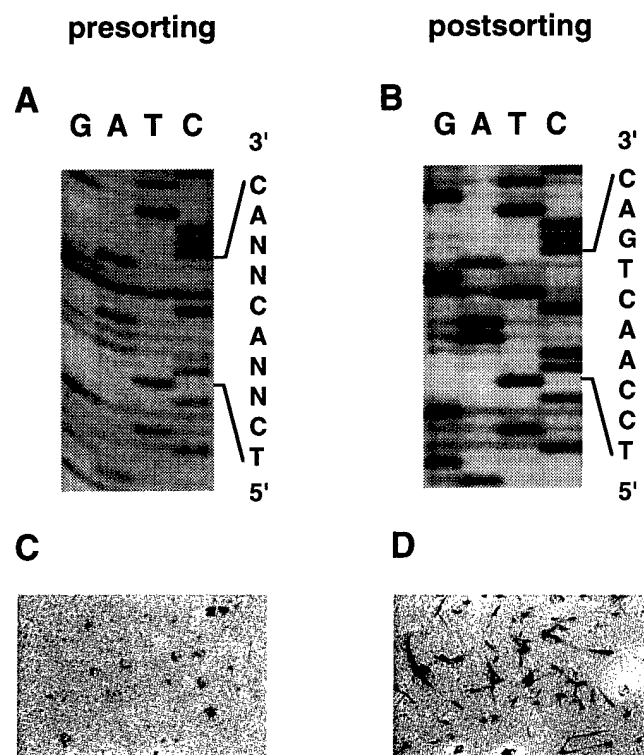


FIG. 2. 4N selection by SITE. (A and B) DNA sequencing gels showing random sequences at desired positions in the initial 4N plasmid pool (A) and emergence of CA and TG dinucleotides from the background after SITE (B). (C and D) X-Gal staining of cells cotransfected with the 4N template and MyoD expression vector before (C) and after (D) SITE.

222N selection, individual bacterial transformants were picked arbitrarily and plasmids were extracted (Miniprep; Promega).

**Sequencing of recovered plasmids.** DNA sequencing was done by the dideoxy method (61), using Sequenase (U.S. Biochemical).

**$\beta$ -Gal assays.**  $\beta$ -Gal activities of transfected cells were monitored by either 5-bromo-4-chloro-3-indolylphosphate- $\beta$ -D-galactopyranoside (X-Gal) staining of fixed cells (60) or *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG) and Galacton-Plus (Tropix) hydrolysis by cell lysates, in accordance with standard protocols (59) or the manufacturer's instructions, respectively.

## RESULTS AND DISCUSSION

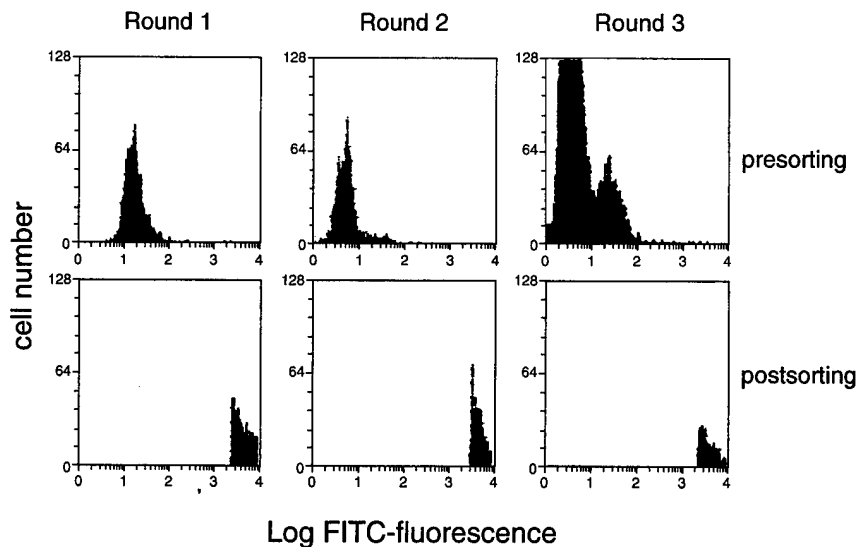
**The CANNTG motif is selected by MyoD to activate transcription.** A transient-transfection approach was used both to generate MyoD-expressing cells and to select MyoD-responsive sequences. This strategy was chosen because stable expression of MyoD inhibits the cell cycle (15, 27, 28, 66) and because myoblast cells that constitutively express MyoD may fuse, thus complicating selection of individual cells. The selection templates (Fig. 1B) were designed on the basis of the HCA promoter (46), which is specific to skeletal and cardiac muscle and contains a single E box that is required for MyoD-dependent HCA promoter activity in skeletal muscle (62, 65). To test the feasibility and fidelity of the SITE approach, we first used the 4N promoter template (Fig. 1B), in which the CA and TG positions of the E box are random. When NIH 3T3 cells were cotransfected with the 4N pool and a MyoD expression vector,  $\beta$ -Gal activity was close to the basal level (Fig. 2C), suggesting that only a very small portion (if any) of the random-sequence combinations were activated by MyoD. In contrast, after two rounds of selection, the pool  $\beta$ -Gal activity was dramatically increased (Fig. 2D), and the DNA sequence of the recovered

plasmids revealed selection for the CA and TG dinucleotides (compare Fig. 2B with Fig. 2A), thus confirming both their importance and the validity of the selection system.

**DNA sequence requirement for MyoD transactivation through the CANNTG motif.** To identify E boxes through which MyoD can activate transcription, we used the 222N template (Fig. 1B), in which base pairs of the HCA promoter internal to and flanking the CANNTG consensus are of random sequence. In preliminary studies, we determined by quantitative  $\beta$ -Gal assay (59) that the wild-type HCA sequence can generate signals at least 50-fold over the level in nontransfected cells or cells transfected with a mutant construct. Consequently, we set the FACS gates for FITC-positive cells at levels well above basal levels in order to allow selection of cells with only the highest levels of  $\beta$ -Gal activity, which probably contain multiple copies of the most active sequences. In the first round, the cells that were cotransfected with the 222N population and MyoD turned the solutions slightly yellow (as a result of FITC leakage) during the loading of FDG, suggesting that the 222N pool contained a considerable portion of active sequence. However, this activity was clearly much lower than that observed for the wild-type sequence. Three rounds of increasingly stringent FACS selection yielded a sequence pool with activity that was significantly higher than that of the starting population, as is apparent from the appearance of a new peak in the FACS spectrum (Fig. 3A). This dramatic improvement in pool activity is also indicated by much more leaking of visible FITC from the cells during the FDG loading procedure in the final round.

When individual plasmids recovered from this pool were assayed for transcriptional activation by cotransfected MyoD, most (15 of 17) were active, some (4 of 17) even more so than the wild-type HCA promoter. Of this last most active group, two (sequences 5 and 15) were represented multiple times (Fig. 3B). None of these selected sequences were active without cotransfection of the MyoD expression vector or could be activated by expression of the E proteins alone. Although a background of inactive plasmids (2 of 26) was still present in the selected pool, these results confirm that the selection procedure enriched significantly for MyoD-responsive sequences (Fig. 3B).

**Differences between in vivo activation and in vitro DNA binding.** Only a subset of the possible E-box CANNTG hexamer core site motifs were represented among the selected molecules that were responsive to MyoD, none of which contained the CACGTG hexamers to which MyoD or MyoD-E complexes are known not to bind (7). Of the 10 sequences that were at least half as active as the wild-type HCA promoter, one contained the wild-type CAAGTG hexamer core sequence, four contained a CAAGTG hexamer, and five contained a CAGGTG sequence. Previously, binding-site selections performed in vitro demonstrated that MyoD homodimers and MyoD-E47 heterodimers prefer sites that contain the CAGCTG and CAGGTG motifs, respectively, and identified their preferred sequences at positions flanking these elements (Fig. 3D) (8). It is remarkable that only a subset of the sequences selected here in vivo contained a CAGGTG motif and that relatively few (4 of 15) of the active plasmids contained the base preferred by MyoD at position  $-4$  or  $+4$ . However, the majority of the sequences selected by this method are identical or very similar to E boxes in the native promoters of the transcriptional regulators MyoD, myogenin, and MRF4, as well as in the promoters of many muscle structural protein genes. For instance, the hexamer core CAAGTG, which was not identified through in vitro binding selection but was recovered repeatedly by SITE, has been observed to date only in the

**A****B**

LABEL	FREQUENCY	SEQUENCE	TRANSCRIPTIONAL ACTIVATION	ANECDOTE
#26	1	GTCAGGTGTG		mAChR-γ R & HCA half-sites
#5	8	ATCAAGTGCA		MyoD DRR1
#18	1	ACCACCTGCA		Desmin ME3 half-site
#15	3	ATCAAGTGCT		MyoD PRR, Myogenin E1 half-site
#40	1	GGCAGGTGTC		MyoD DRR3, MCK R
#28	1	AGCAGGTGTG		MCK R, sTnC, MLC 1/3A, Myogenin E1
#27	1	TGCAAGTGAG		MyoD PRR & DRR1 core
#9	1	GGCAGTTGAA		Myogenin E1, HCA
#30	1	GGCAACTGGA		MRF4 E2
#36	1	GTCAGGTGGC		mAChR-γ R & HCA half-sites
#8	1	GACAGATGGG		mAChR-γ L, MRF4 E1, AchR-δ E1, Desmin ME3
#17	1	GTCAAGTGGG		MyoD PRR
#10	1	CCCACATGTG		MCK L
#19	1	GTCAGGTGAG		mAChR-γ R, HCA & #27 half-sites
#29	1	ATCAGGTGAA		both half-sites unfavorable for E
#6	1	AACAATTGGT		background
#22	1	CTCATTTGAA		background
w.t.	N/A	TCCAAGTGAC		w.t. HCA
mut	N/A	TTGGTCCTGA		mutant HCA
gc	N/A	GTCAGCTGAC		in vitro MyoD homodimer site
aa	N/A	AACAGGTGTG		in vitro optimal M-E binding site

**C**

(a) #5	ATCAAGTGCA
MyoD DRR1	ACCAAGTGTA
(b) #15	ATCAAGTGCT
MyoD PRR	GGCAAGTGCT
Myogenin E1	GGCAAGTGCT (-)
(c) #17	GTCAAGTGGG
MyoD PRR	GGCAAGTGCT
#27	TGCAAGTGAG
(d) #26	GTCAGGTGTG
AChR-γ R	GTCAGCTGTT
HCA	GTCAGTTGGA
#36	GTCAGGTGGC
(e) #18	TGCAGGTGCT
Desmin E3	TGCAGGTGTC
(f) #40	GGCAGGTGTC
MyoD DRR3	GGCAGGTGCT
MCK R	AGCAGGTGTT (-)
Desmin E2	GGCAGCTGTT

(g) #28	AGCAGGTGTG
MLC 1/3A	AGCAGGTGCA
MCK R	AGCAGGTGTT (-)
sTnC	AGCAGGTG (-)
Myogenin E1	AGCAGTTGGC
(h) #8	GACAGATGGG
mAChR-γ L	GACAGATGGG (-)
MRF4 E1	CCCAGATGGC (-)
MyoD DRR2	GACAGGTGGA (-)
AChR-δ E1	GACAGGTGCT (-)
cAChR-γ R	GACAGCTGTC
Desmin E3	GACAGCTGCA (-)
qTnI IRE	GACAGCTGCA
(i) #9	GGCAGTTGAA
Myogenin E1	AGCAGTTGGC
HCA	GTCAGTTGGA
(j) #30	TCCAAGTGGC (-)
MRF4 E2	CCCAGTTGTT
(k) #10	CCCACATGTG
MCK L	CCCACATGTC (-)

D MyoD*	AACAGCTGTT
Gt	aC
Myogenin†	AACAGCTGTT
	T
E47*	NTCACCTGAa
E47-MyoD*	NTCACCTGTN
E12-MyoD*	CTCACCTGTT
	C
E-myogenin†	TGCACCTGTT



FIG. 3. SITE-selected sequences and transcriptional activities from the 222N template. (A) FACS sorting of the 222N pool. The sort gates for  $\beta$ -Gal activity were set such that the ratios of FITC fluorescence intensity of the gate over that of the bulk of  $\beta$ -Gal-negative cells were 150, 500, and 700 for the first, second, and third rounds, respectively. Note that the fluorescence intensity axis is logarithmic. Dead cells (propidium iodide bright) were eliminated through a second fluorescence gating. Cell debris, doublets, and larger clumps were gated out by forward- and side-scattering settings. The upper panels are histograms of the FITC intensities of all cells; the lower panels show the sorted cells while they were accumulating over an arbitrary period during the sorting. The cells containing FITC levels higher than the sort gate were collected, and the plasmid DNA was extracted and amplified in *E. coli* cells in order to do the next round of SITE. (B) Transcriptional activation by sequences selected from the third round of SITE. The plasmids recovered from the third round of the selection were transfected individually with the MyoD expression vector into 3T3 cells.  $\beta$ -Gal activity was measured from the cell lysates, using ONPG as the substrate. Each datum point was the average of at least three independent transfections. The data were plotted as the fold activation by MyoD over that of pEMSVscribe (MyoD<sup>-</sup> control vector). The Anecdote column lists known muscle-specific enhancer/promoters that contain E boxes with sequences identical or very similar to those selected by SITE. Only E boxes that were tested to be functional in the literature are listed. Note that in sequence 5, although the 3' flanking sequence CA can in theory be grouped with nucleotides immediately downstream to form another E box (CACCTG site), this second E box is nonfunctional (data not shown). mAChR- $\gamma$ , mouse acetylcholine receptor  $\gamma$  subunit; DRR1, distal regulatory region 1; PRR, proximal regulatory region; sTnC, smooth muscle troponin C; MLC, myosin light chain; w.t., wild type; mut, mutant. (C) Alignment of selected E-box sequences to known enhancer/promoter sequences in muscle-specific genes. (—), the sequence displayed is antisense to the template strand. (D) Summary of in vitro DNA binding preferences of myogenic bHLH and E proteins. \*, data from reference 8; †, data from reference 24; ‡, data from reference 80. A line drawn over an uppercase letter indicates a base that is absent in that position, and a line over a lowercase letter indicates a decrease in use of that base.

MyoD promoter. Among them, sequence 5 resembles the MyoD distal regulatory region 1, which is important for transactivation of the MyoD promoter (69), and sequence 15 is very similar to the E box in the MyoD proximal regulatory region, which is indispensable for both distal regulatory region transactivation (69) and autoregulation (80) by MyoD. Although the disproportionate representation of sequences 5 and 15 suggests that the arbitrarily chosen number of cells from which plasmids were recovered may have contained sequences of limited complexity, a variety of different sequences that indeed differ from the preferred in vitro consensus were selected.

Perhaps the most striking aspect of the selected MyoD-responsive sequences is that they closely resemble functional E boxes that are present in the promoters of muscle-specific genes (Fig. 3B and C) (5, 13, 14, 23, 26, 38, 40, 42, 43, 57, 63, 78) but not those found in the target genes of nonmuscle bHLH proteins (see references in reference 8). This resemblance is apparent not only in their core hexamers but also in flanking sequences, many with identical pentamer half-sites (Fig. 3C). In their natural promoter context, they are also capable of mediating transcriptional activation by the myogenic bHLH proteins, which apparently function as heterodimers with E proteins. These muscle-specific sites can, in general, be bound in vitro by MyoD and the other myogenic bHLH proteins and by heterodimers of these proteins with E proteins. The MyoD-responsive sequences selected here by SITE can similarly be bound by MyoD-E heterodimers (although the in vitro binding affinities do not parallel the in vivo transcriptional activity [data not shown]), and when we examined them in the cotransfection assay, we found that they mediate activation by a tethered MyoD-E47 dimer (52) to levels no higher than those characteristic of cotransfected MyoD (data not shown). The simplest interpretation of these findings is that these selected sites are direct targets of heterodimers of cellular E proteins with the cotransfected MyoD or with other myogenic bHLH proteins that may be activated by MyoD (see below).

The SITE data thus support the notion that DNA binding alone is not sufficient for a bound MyoD (or other myogenic bHLH protein)-E heterodimer to be transcriptionally active. To confirm that this SITE experiment did not somehow simply miss functional sequences that more closely resemble optimal MyoD and MyoD-E in vitro binding sites, we tested such sites (Fig. 3B) in the HCA cassette for transactivation by MyoD. In this assay, surprisingly, an optimal in vitro MyoD-E site was inactive, and the activity of a preferred MyoD homodimer site was very low (Fig. 3B). On the other hand, SITE recovered some transcriptionally active plasmids bearing a sequence that was strongly selected against by in vitro methods, e.g., T at the -4 position of the E47 protein half-site (8, 67). These data

indicate that the E boxes that are most responsive to MyoD in vivo do not necessarily correspond to the in vitro binding preferences of MyoD or MyoD-E complexes.

**Role of promoter context in MyoD transactivation through different E boxes.** One possible explanation for differences in vivo and in vitro is that in vivo repressors might bind to certain E-box sequences and prevent MyoD from binding to these sites and activating transcription. Alternatively, differences among interacting proteins on these promoters might affect the conformation of the DNA and/or of the protein-DNA complex in a way that prevents particular E-box sequences from acting as an effective enhancer. To distinguish between these possibilities, we compared the abilities of different E boxes to activate transcription in the HCA promoter and the MCK enhancer/promoter. The 3,300-bp enhancer/promoter of MCK contains two E boxes that are targets of MyoD activation (37, 74). A 115-bp segment including these E boxes that retains all of the muscle-specific activity of the enhancer (35) was used for the E-box substitution experiment (Fig. 4A), the results of which are shown in Fig. 4B. The sequence 5 E-box decamer, a strong binding site in vitro and a strong activation site in the HCA promoter, was only moderately active when replacing the

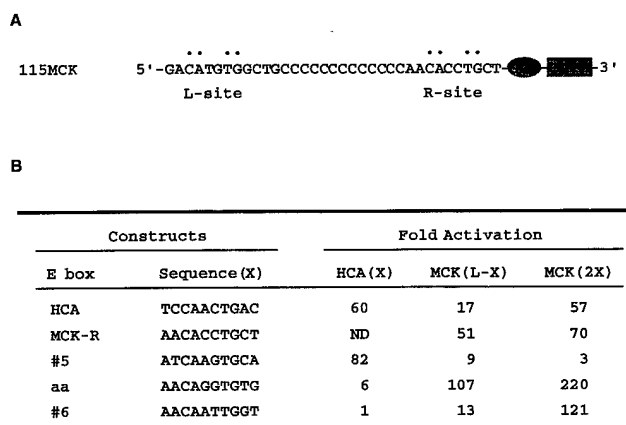


FIG. 4. Comparison of transcriptional activities of different E boxes in the MCK and HCA promoter contexts. (A) Molecular structure and enhancer sequence of the MCK enhancer and promoter. Only the two E boxes (L and R sites) are shown. (B) Comparison of transcriptional activities of different E boxes in the MCK and HCA promoter contexts. X's stand for different E-box sequences; HCA (X) denotes X replacing the E box in the HCA promoter, MCK (L-X) denotes X replacing the R site in the MCK enhancer, and MCK (2X) denotes X replacing both the L and R sites in the MCK enhancer. Each datum point was derived from transfection of two independent clones.  $\beta$ -Gal activity was measured from cell lysates by using Galacton-Plus (Tropix) as the substrate.

right (R) site of the MCK enhancer. Its activity was even lower if both left (L) and R sites were replaced. On the other hand, the in vitro optimal binding site (construct aa), although inactive in the HCA context, was the most active in the MCK context. It is interesting that the HCA E box, a poor binding site in vitro, acts as a strong activation site both in its native context and when replacing the L and R sites simultaneously in the MCK enhancer. Thus, these data for the MCK promoter confirmed the notion that the ability of an E-box enhancer to support transcriptional activation cannot be predicted from its affinity for the protein factors. These results demonstrate that promoter context is an important determinant of activation properties of different E-box sequences.

It is interesting that a majority (20 of 26) of the sites that were selected bear a MyoD half-site on the 5' end in proximity to the Sp1 site on the HCA promoter (Fig. 1B and 3B). However, it is unlikely that Sp1 could be the determinant of context specificity because the MCK enhancer also contains an Sp1 site. Sp1 sites are also found adjacent to MyoD half-sites in the promoters of the acetylcholine receptor  $\delta$  subunit (AChR- $\delta$ ), myogenin, troponin, and MyoD genes, although the importance of these Sp1 sites has not been uniformly evaluated. Thus, it is not obvious why the activities of some E boxes are so different in the HCA and MCK promoters. Our results thus form a basis for further investigation of *cis*-acting elements that are involved in determining promoter context specificity.

**Application of the SITE method.** The SITE technique described here has, for the first time, used the power of random-sequence selection (6, 54, 64, 68) in metazoan cells to identify functional regulatory sites. It can yield sequences that are responsive to any transcription factor, or to a protein that activates such a factor, provided that either the cloned gene or a cell line expressing the factor is available. The reiterative selection process permits recovery of a large number of sites that mediate a range of transcriptional activities, with more stringent selection yielding more active sites. This method is likely to yield biologically significant sequences that might not be identified by in vitro methods, because they are selected in the presence of the full complement of cellular factors under physiological conditions and can thus provide rational guidance for designing reporter constructs. As such, the in vivo selection method provides an approach complementary to the in vitro method for discovering target sequences for regulatory proteins. A more important benefit of this strategy is that it allows a systematic comparison of the functional requirements of proteins under different circumstances and conditions and can thus illuminate critical regulatory mechanisms. For example, a comparison of the MyoD-responsive sequences identified in this study with the preferred in vitro binding sites of MyoD suggests the existence of particular sequence requirements for activation. The versatility of the SITE approach will now make it possible to undertake similar investigations of activation by MyoD (and its BR variants) in the context of other promoters and associated factors and to identify similarly sequences that respond to other bHLH proteins.

The SITE technique should be adaptable to a variety of other experimental situations. For example, repression sequences could be selected, provided that a screening step first identified cells that contain the reporter library DNA. It should be similarly possible to select native regulatory sequences from a library of genomic DNA fragments or, if transcribed positions were randomized, to identify sequences involved in RNA processing or translational control. If positions within coding sequences were randomized, a selection could identify the amino acid residues that are encoded by the corresponding positions which allow a transcription factor to function, as has

been done in bacteria (34) and in *Saccharomyces cerevisiae* (56). In principle, the number of positions that can be randomized in the SITE technique should be limited only by the number of cells that can be handled. For example, eight random positions will generate  $4^8$  (about  $6.6 \times 10^4$ ) different plasmid molecules. If one molecule were introduced per cell, screening of a library of this size would require on the order of  $3 \times 10^5$  cells (59), a number readily attainable by transient transfection. However, the efficiency of screening is actually higher, because each cell takes up multiple plasmids and because even at a high input copy number, positive cells can be identified by FACS, which can detect as few as 5  $\beta$ -Gal enzyme molecules per cell (53), is rapid (1/10<sup>6</sup> cells per h), and can be applied to different markers (e.g., the cell surface molecule CD4 [45] or green fluorescent protein [GFP] [11]). Multiple markers can be used either alone or simultaneously to examine different sets of regulatory sequences. Alternatively, the technique can be applicable to other transient selection markers (e.g., puromycin resistance [71]).

**Specificity of MyoD-mediated transcriptional activation.** Although the sequences identified here are responsive to MyoD, it is possible that these sequences actually are or include the direct targets of another myogenic bHLH protein. In cultured cells, expression of a myogenic bHLH protein can activate the corresponding endogenous gene along with other members of this group, with the particular myogenic genes activated varying among different lines (2, 33, 70). In the 3T3 cells used in this study, MyoD generally does not activate the endogenous MyoD, myf-5, or MRF4 gene but can activate the myogenin gene. In the mouse, myf-5 and MyoD determine the myoblast cell fate (58), while myogenin is required to activate the full spectrum of sarcomeric genes (30, 50), but in cells cultured from myogenin<sup>-/-</sup> mice, MyoD can also activate these genes (50). Accordingly, we cannot determine whether in our experiments myogenin might have also contributed to the activation of sequences selected. However, it is unlikely that only myogenin (but not MyoD) is the direct activator, because when the selected sequences were cotransfected with a myogenin expression vector, they were activated to levels only half as high as those achieved by MyoD. As these sites are bound by MyoD-E heterodimers in vitro and activated similarly by MyoD and MyoD-E tethered dimers in vivo (data not shown), and since the DNA binding requirements of MyoD in vitro and those of myogenin with the nuclear extracts are virtually identical (Fig. 3D) (8, 79), the most likely conclusion appears to be that the selected regulatory sites may be the direct targets of heterodimers of either MyoD or myogenin with endogenous cellular E proteins.

The group of sequences that were identified in this experiment share some attributes with, but are not identical to, the preferred in vitro binding sites for MyoD (or myogenin) or MyoD-E complexes. In the HCA promoter, these bHLH complexes might bind cooperatively with neighboring proteins, and these interactions might have contributed enough binding energy to have relaxed the sequence requirements for activation. However, two lines of evidence argue against this interpretation. First, an optimal site was not capable of supporting activation in this context (Fig. 3B). Second, the selected sequences were actually less similar to the in vitro preferences than to E boxes found in the native promoters of muscle-specific genes. These native E boxes have been shown to be important for auto- and/or cross-activation of the myogenic bHLH genes or for activation of muscle structural genes by these bHLH proteins, suggesting that the sequences selected here by SITE are biologically significant.

We have selected E-box sequences that are able to support

transcriptional activation by muscle-specific bHLH protein MyoD. The data indicate that only a subset of MyoD binding sites allow activation of transcription in the HCA promoter context and that certain E boxes function differently in different muscle-specific promoters. These findings demonstrate that promoter context plays an important role in transcriptional activation by MyoD through E boxes. Promoter context could affect transcriptional activation in several ways. For example, the neighboring protein factors could influence the conformation of the promoter DNA and/or of the protein-E box complex; different proteins could also recruit distinct associated factors to the promoters. In addition, our findings are reminiscent of the positive-control MyoD BR mutations (17, 76), which also appear to permit DNA binding but not activation. Significantly, these same BR residues not only appear to influence the conformation of the protein-DNA complex (19, 44) but also have been implicated in interactions with proposed regulatory factors (9, 16, 17, 47, 76). The findings described here indicate that some, or perhaps all, of the potential regulatory interactions require specific DNA sequences, which themselves could influence the conformation of the bound MyoD-E complex. It remains to be determined whether such conformational changes might provide distinctive targets for coactivators or repressors or would derive from such interactions.

#### ACKNOWLEDGMENTS

We thank members of the Weintraub and Kedes laboratories for advice and support; S. Fiering for initial help and advice with FACS; A. Berger for assistance with numerous cell sortings; V. Sartorelli for the HCA promoter; B. Wold for the MyoD-E47 expression vector; A. Jimenez and A. Lassar for pac cDNA and expression vector; M. Emerman for CD4 cDNA; M. Chalfie for GFP cDNA; N. Rosenthal for sharing information; and D. Turner, S. Handeli, J. Gogos, Y. Zhuang, M. Horwitz, K. Munir, and R. Schreck for helpful discussions and encouragement. J.H. is grateful to J. T. Huang, M. H. Wu, F. Zhang, S. Lew, H.W., T.K.B., and L.K. for faith, understanding, and support. We thank M. Groudine, H. Sucov, and B. Wold for critical reading of the manuscript.

This work was supported in part by grants from NIH (to H.W. and to L.K.), the U.S. Army Breast Cancer Research Program (to T.K.B.), and HHMI (to H.W.).

#### REFERENCES

- Aplan, P. D., K. Nakahara, S. H. Orkin, and I. R. Kirsch. 1992. The SCL gene product: a positive regulator of erythroid differentiation. *EMBO J.* 11:4073-4081.
- Aurade, F., C. Pinset, P. Chafey, F. Gros, and D. Montarras. 1994. Myf5, MyoD, myogenin and MRF4 myogenic derivatives of the embryonic mesenchymal cell line C3H10T1/2 exhibit the same adult muscle phenotype. *Differentiation* 55:185-192.
- Bain, G., S. Gruenwald, and C. Murre. 1993. E2A and E2-2 are subunits of B-cell-specific E2-box DNA-binding proteins. *Mol. Cell. Biol.* 13:3522-3529.
- Benezra, R. 1994. An intermolecular disulfide bond stabilizes E2A homodimers and is required for DNA binding at physiological temperatures. *Cell* 79:1057-1067.
- Black, B. L., J. F. Martin, and E. N. Olson. 1995. The mouse MRF4 promoter is trans-activated directly and indirectly by muscle-specific transcription factors. *J. Biol. Chem.* 270:2889-2892.
- Blackwell, T. K. 1995. Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol.* 254:608-618.
- Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman, and H. Weintraub. 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* 13:5216-5224.
- Blackwell, T. K., and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* 250:1104-1110.
- Brennan, T. J., T. Chakraborty, and E. N. Olson. 1991. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl. Acad. Sci. USA* 88:5675-5679.
- Buckingham, M. E. 1994. Muscle: the regulation of myogenesis. *Curr. Opin. Genet. Dev.* 4:745-751.
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263:802-805.
- Chen, C., and H. Okayama. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* 7:2745-2752.
- Cheng, T. C., M. C. Wallace, J. P. Merlie, and E. N. Olson. 1993. Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science* 261:215-218.
- Christensen, T. H., H. Prentice, R. Gahlmann, and L. Kedes. 1993. Regulation of the human cardiac/slow-twitch troponin C gene by multiple, cooperative, cell-type-specific, and MyoD-responsive elements. *Mol. Cell. Biol.* 13:6752-6765.
- Crescenzi, M., T. P. Fleming, A. B. Lassar, H. Weintraub, and S. A. Aaronson. 1990. MyoD induces growth arrest independent of differentiation in normal and transformed cells. *Proc. Natl. Acad. Sci. USA* 87:8442-8446.
- Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* 60:733-746.
- Davis, R. L., and H. Weintraub. 1992. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* 256:1027-1030.
- Desiderio, S. 1995. Transcription factors controlling B-cell development. *Curr. Biol.* 5:605-608.
- Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8:970-980.
- Emerson, C. P. 1990. Myogenesis and developmental control genes. *Curr. Opin. Cell Biol.* 2:1065-1075.
- Ferre-D'Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley. 1993. Recognition by Max of its cognate DNA through a dimeric bHLH/Z domain. *Nature (London)* 363:38-45.
- Fiering, S. N., M. Roederer, G. P. Nolan, D. R. Micklem, D. R. Parks, and L. A. Herzenberg. 1991. Improved FACS-Gal: flow cytometric analysis and sorting of viable eukaryotic cells expressing reporter gene constructs. *Cytometry* 12:291-301.
- Fujisawa, S. A., K. Hanaoka, M. Hayasaka, Y. T. Hiromasa, and Y. Nabeshima. 1993. Upstream region of the myogenin gene confers transcriptional activation in muscle cell lineages during mouse embryogenesis. *Biochem. Biophys. Res. Commun.* 191:351-356.
- Funk, W. D., and W. E. Wright. 1992. Cyclic amplification and selection of targets for multicomponent complexes: myogenin interacts with factors recognizing binding sites for basic helix-loop-helix, nuclear factor 1, myocyte-specific enhancer-binding factor 2, and COM1 factor. *Proc. Natl. Acad. Sci. USA* 89:9484-9488.
- Genetta, T., D. Ruzinsky, and T. Kadesch. 1994. Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.* 14:6153-6163.
- Gilmour, B. P., G. R. Fanger, C. Newton, S. M. Evans, and P. D. Gardner. 1991. Multiple binding sites for myogenic regulatory factors are required for expression of the acetylcholine receptor gamma-subunit gene. *J. Biol. Chem.* 266:19871-19874.
- Gu, W., J. W. Schneider, G. Condorelli, S. Kaushal, V. Mahdavi, and G. B. Nadal. 1993. Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation. *Cell* 72:309-324.
- Halevy, O., B. G. Novitch, D. B. Spicer, S. X. Skapek, J. Rhee, G. J. Hannon, D. Beach, and A. B. Lassar. 1995. Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD. *Science* 267:1018-1021.
- Hanahan, D., J. Jessee, and F. R. Bloom. 1991. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* 204:63-113.
- Hasty, P., A. Bradley, J. H. Morris, D. G. Edmondson, J. M. Venuti, E. N. Olson, and W. H. Klein. 1993. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature (London)* 364:501-506.
- Higuchi, R. 1989. Using PCR to engineer DNA, p. 61-70. *In* H. A. Erlich (ed.), PCR technology. Stockton Press, New York.
- Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* 26:365-369.
- Hollenberg, S. M., P. F. Cheng, and H. Weintraub. 1993. Use of a conditional MyoD transcription factor in studies of MyoD trans-activation and muscle determination. *Proc. Natl. Acad. Sci. USA* 90:8028-8032.
- Hu, J. C., E. K. O'Shea, P. S. Kim, and R. T. Sauer. 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* 250:1400-1403.
- Huang, J., and H. Weintraub. Unpublished data.
- Jan, Y. N., and L. Y. Jan. 1993. HLH proteins, fly neurogenesis, and vertebrate myogenesis. *Cell* 75:827-830.
- Jaynes, J. B., J. E. Johnson, J. N. Buskin, C. L. Gartside, and S. D. Hauschka. 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. *Mol. Cell. Biol.* 8:62-70.
- Jia, H. T., H. J. Tsay, and J. Schmidt. 1992. Analysis of binding and acti-

- vating functions of the chick muscle acetylcholine receptor gamma-subunit upstream sequence. *Cell. Mol. Neurobiol.* **12**:241-258.
39. Kadesch, T. 1992. Helix-loop-helix proteins in the regulation of immunoglobulin gene transcription. *Immunol. Today* **13**:31-36.
  40. Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub. 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* **58**:823-831.
  41. Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66**:305-315.
  42. Li, H., and Y. Capetanaki. 1994. An E box in the desmin promoter cooperates with the E box and MEF-2 sites of a distal enhancer to direct muscle-specific transcription. *EMBO J.* **13**:3580-3589.
  43. Lin, H., K. E. Yutzy, and S. F. Konieczny. 1991. Muscle-specific expression of the troponin I gene requires interactions between helix-loop-helix muscle regulatory factors and ubiquitous transcription factors. *Mol. Cell. Biol.* **11**:267-280.
  44. Ma, P. C., M. A. Rould, H. Weintraub, and C. O. Pabo. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**:451-459.
  45. Maddon, P. J., D. R. Littman, M. Godfrey, D. E. Maddon, L. Chess, and R. Axel. 1985. The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. *Cell* **42**:93-104.
  46. Minty, A., and L. Kedes. 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. *Mol. Cell. Biol.* **6**:2125-2136.
  47. Molkenkin, J. D., B. L. Black, J. F. Martin, and E. N. Olson. 1995. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* **83**:1125-1136.
  48. Murre, C., P. S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**:777-783.
  49. Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**:537-544.
  50. Nabeshima, Y., K. Hanaoka, M. Hayasaka, E. Esumi, S. Li, I. Nonaka, and Y. Nabeshima. 1993. Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature (London)* **364**:532-535.
  51. Nelson, C., L. P. Shen, A. Meister, E. Fodor, and W. J. Rutter. 1990. Pan: a transcriptional regulator that binds chymotrypsin, insulin, and AP-4 enhancer motifs. *Genes Dev.* **4**:1035-1043.
  52. Neuhold, L. A., and B. Wold. 1993. HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* **74**:1033-1042.
  53. Nolan, G. P., S. Fiering, J. F. Nicolas, and L. A. Herzenberg. 1988. Fluorescence-activated cell analysis and sorting of viable mammalian cells based on beta-D-galactosidase activity after transduction of *Escherichia coli* lacZ. *Proc. Natl. Acad. Sci. USA* **85**:2603-2607.
  54. Oliphant, A. R., and K. Struhl. 1988. Defining the consensus sequences of *E. coli* promoter elements by random selection. *Nucleic Acids Res.* **16**:7673-7683.
  55. Olson, E. N. 1992. Proto-oncogenes in the regulatory circuit for myogenesis. *Semin. Cell Biol.* **3**:127-136.
  56. Pu, W. T., and K. Struhl. 1993. Dimerization of leucine zippers analyzed by random selection. *Nucleic Acids Res.* **21**:4348-4355.
  57. Rosenthal, N., E. B. Berglund, B. M. Wentworth, M. Donoghue, B. Winter, E. Bober, T. Braun, and H. H. Arnold. 1990. A highly conserved enhancer downstream of the human MLC1/3 locus is a target for multiple myogenic determination factors. *Nucleic Acids Res.* **18**:6239-6246.
  58. Rudnicki, M. A., P. N. Schnegelsberg, R. H. Stead, T. Braun, H. H. Arnold, and R. Jaenisch. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**:1351-1359.
  59. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  60. Sanes, J. R., J. L. Rubenstein, and J. F. Nicolas. 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. *EMBO J.* **5**:3133-3142.
  61. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
  62. Sartorelli, V., K. A. Webster, and L. Kedes. 1990. Muscle-specific expression of the cardiac alpha-actin gene requires MyoD1, CArG-box binding factor, and Sp1. *Genes Dev.* **4**:1811-1822.
  63. Simon, A. M., and S. J. Burden. 1993. An E box mediates activation and repression of the acetylcholine receptor  $\delta$ -subunit gene during myogenesis. *Mol. Cell. Biol.* **13**:5133-5140.
  64. Singer, V. L., C. R. Wobbe, and K. Struhl. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev.* **4**:636-645.
  65. Skerjanc, I. S., and M. W. McBurney. 1994. The E box is essential for activity of the cardiac actin promoter in skeletal but not in cardiac muscle. *Dev. Biol.* **163**:125-132.
  66. Sorrentino, V., R. Pepperkok, R. L. Davis, W. Ansorge, and L. Philipson. 1990. Cell proliferation inhibited by MyoD1 independently of myogenic differentiation. *Nature (London)* **345**:813-815.
  67. Sun, X. H., and D. Baltimore. 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* **64**:459-470.
  68. Szostak, J. W. 1992. In vitro genetics. *Trends Biochem. Sci.* **17**:89-93.
  69. Tapscott, S. J., A. B. Lassar, and H. Weintraub. 1992. A novel myoblast enhancer element mediates MyoD transcription. *Mol. Cell. Biol.* **12**:4994-5003.
  70. Thayer, M. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub. 1989. Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**:241-248.
  71. Vara, J., F. Malpartida, D. A. Hopwood, and A. Jimenez. 1985. Cloning and expression of a puromycin N-acetyl transferase gene from *Streptomyces alboniger* in *Streptomyces lividans* and *Escherichia coli*. *Gene* **33**:197-206.
  72. Voronova, A., and D. Baltimore. 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* **87**:4722-4726.
  73. Weintraub, H. 1993. The MyoD family and myogenesis: redundancy, networks, and thresholds. *Cell* **75**:1241-1244.
  74. Weintraub, H., R. Davis, D. Lockshon, and A. Lassar. 1990. MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. *Proc. Natl. Acad. Sci. USA* **87**:5623-5627.
  75. Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, and S. Hollenberg. 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**:761-766.
  76. Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott. 1991. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **5**:1377-1386.
  77. Weintraub, H., T. Genetta, and T. Kadesch. 1994. Tissue-specific gene activation by MyoD: determination of specificity by cis-acting repression elements. *Genes Dev.* **8**:2203-2211.
  78. Wentworth, B. M., M. Donoghue, J. C. Engert, E. B. Berglund, and N. Rosenthal. 1991. Paired MyoD-binding sites regulate myosin light chain gene expression. *Proc. Natl. Acad. Sci. USA* **88**:1242-1246.
  79. Wright, W. E., M. Binder, and W. Funk. 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol. Cell. Biol.* **11**:4104-4110.
  80. Zingg, J. M., A. G. Pedraza, and J. P. Jost. 1994. MyoD1 promoter autoregulation is mediated by two proximal E-boxes. *Nucleic Acids Res.* **22**:2234-2241.

# Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression

Jitakshi De <sup>a</sup>, Wi S. Lai <sup>a</sup>, Judith M. Thorn <sup>a</sup>, Susan M. Goldsworthy <sup>a</sup>, Xiaoying Liu <sup>b</sup>,  
T. Keith Blackwell <sup>b</sup>, Perry J. Blackshear <sup>a,c,\*</sup>

<sup>a</sup> Office of Clinical Research and Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

<sup>b</sup> Center for Blood Research, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA

<sup>c</sup> Departments of Medicine and Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

Received 12 October 1998; received in revised form 8 December 1998; accepted 9 December 1998, Received by A. Dugaiczky

## Abstract

Tristetraprolin (TTP), the prototype of a class of CCCH zinc finger proteins, is a phosphoprotein that is rapidly and transiently induced by growth factors and serum in fibroblasts. Recent evidence suggests that a physiological function of TTP is to inhibit tumor necrosis factor  $\alpha$  secretion from macrophages by binding to and destabilizing its mRNA (Carballo, E., Lai, W.S., Blackshear, P.J., 1998. *Science*, 281, 1001–1005). To investigate possible functions of CCCH proteins in early development of *Xenopus*, we isolated four *Xenopus* cDNAs encoding members of this class. Based on 49% overall amino acid identity and 84% amino acid identity within the double zinc finger domain, one of the *Xenopus* proteins (XC3H-1) appears to be the homologue of TTP. By similar analyses, XC3H-2 and XC3H-3 are homologues of ERF-1 (cMG1, TIS11B) and ERF-2 (TIS11D). A fourth protein, XC3H-4, is a previously unidentified member of the CCCH class of vertebrate zinc finger proteins; it contains four C<sub>x</sub><sub>8</sub>C<sub>x</sub><sub>5</sub>C<sub>x</sub><sub>3</sub>H repeats, two of which are YKTEL C<sub>x</sub><sub>8</sub>C<sub>x</sub><sub>5</sub>C<sub>x</sub><sub>3</sub>H repeats that are closely related to sequences found in the other CCCH proteins. Whereas XC3H-1, XC3H-2, and XC3H-3 were widely expressed in adult tissues, XC3H-4 mRNA was not detected in any of the adult tissues studied except for the ovary. Its expression appeared to be limited to the ovary, oocyte, egg and the early embryonic stages leading up to the mid-blastula transition. Its mRNA was highly expressed in oocytes of all ages, and was enriched in the animal pole cytosol of mature oocytes. Maternal expression was also seen with the other three messages, suggesting the possibility that these proteins are involved in regulating mRNA stability in oocyte maturation and/or early embryogenesis. © 1999 Elsevier Science B.V. All rights reserved.

**Keywords:** AU-rich element; Deadenylation; Mid-blastula transition; mRNA stability

## 1. Introduction

Members of the CCCH class of zinc finger proteins contain two or more zinc finger motifs, each with a cysteine–histidine repeat in a cys–cys–cys–his (CCCH)

configuration. Members of a subclass of this family contain two putative zinc fingers with tandem YKTEL C<sub>x</sub><sub>8</sub>C<sub>x</sub><sub>5</sub>C<sub>x</sub><sub>3</sub>H repeats (where *x* is a variable amino acid), in which the H of the first finger is separated from the first C of the second finger by 18 amino acids. Three members of this group have thus far been identified in mammals: Tristetraprolin (TTP) (Lai et al., 1990), also known as TIS11 (Varnum et al., 1989; Ma and Herschman, 1991), Nup475 (DuBois et al., 1990), and G0S24 (Heximer and Forsdyke, 1993); cMG1 (Gomperts et al., 1990), also known as TIS11B (Varnum et al., 1991), ERF-1 (Barnard et al., 1993), and Berg36 (Ning et al., 1996); and TIS11D (Varnum et al., 1991), also known as ERF-2 (Nie et al., 1995). The structure

\* Corresponding author. A2-05 NIEHS, 111 Alexander Drive, Research Triangle Park, NC 27709, USA. Tel.: +1-919-541-4899; fax: +1-919-541-4571; e-mail: black009@niehs.nih.gov.

Abbreviations: ARE, AU-rich element; bp, base pair(s); CPSF, cleavage and polyadenylation specificity factor; HCG, human chorionic gonadotropin; kb, kilobase(s); MBS, modified Barth's saline; MBT, mid-blastula transition; PABP, poly A binding protein; RACE, rapid amplification of cDNA ends; TNF $\alpha$ , tumor necrosis factor  $\alpha$ ; TTP, tristetraprolin; UTR, untranslated region.

of one of the previously hypothetical zinc fingers has recently been resolved by nuclear magnetic resonance spectroscopy; this finger has been shown to bind zinc with high affinity (Worthington et al., 1996). Other members of this gene family have been identified in organisms as diverse as *Drosophila*, *C. elegans*, and yeast (Mello et al., 1992, 1996; Ma et al., 1994; Warbrick and Glover, 1994; Seydoux et al., 1996; Thompson et al., 1996; Guedes and Priess, 1997).

TTP, the best studied member of this family, derives its name from three PPPPG motifs that are highly conserved in mammals (Taylor et al., 1995). It is encoded by *Zfp36* in mouse and *ZFP36* in man, which have been mapped to chromosomes 7 and 19q13.1 respectively (Taylor et al., 1991). In fibroblasts, it is rapidly induced at the level of transcription by several mitogens, including insulin, serum, platelet-derived growth factor, fibroblast growth factor, and phorbol 12-myristate 13-acetate (Varnum et al., 1989; DuBois et al., 1990; Lai et al., 1990). In addition, TTP is rapidly phosphorylated on serine residues after stimulation of cells by the same mitogens (Taylor et al., 1995), which also rapidly stimulate translocation of TTP from the nucleus to the cytoplasm in fibroblasts (Taylor et al., 1996a,b). Because there is evidence that TTP binds  $Zn^{2+}$  (DuBois et al., 1990; Worthington et al., 1996) and is localized, at least part of the time, in the nucleus (DuBois et al., 1990; Taylor et al., 1996a,b), it has been suggested that TTP may be a nuclear transcription factor like the other immediate-early response genes.

Mice made deficient in TTP by gene targeting appeared normal at birth but soon developed a complex syndrome that included cachexia, dermatitis, erosive arthritis, autoimmunity, and myeloid hyperplasia; essentially all aspects of this phenotype could be prevented by the injection of antibodies to tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) (Taylor et al., 1996a,b). TTP was later found to inhibit TNF $\alpha$  production from macrophages and perhaps other cells, in that macrophages from TTP-knockout mice produced excessive amounts of TNF $\alpha$  (Carballo et al., 1997). This was associated with increased cellular concentrations of TNF $\alpha$  mRNA (Carballo et al., 1997).

The mechanism of this increase in TNF $\alpha$  mRNA levels was recently found to be due, at least in part, to stabilization of TNF $\alpha$  mRNA in the TTP-deficient macrophages (Carballo et al., 1998). This was associated with direct binding of TTP to the AU-rich element (ARE) of the TNF $\alpha$  mRNA (Carballo et al., 1998). Thus, TTP appears to enhance the breakdown of TNF $\alpha$  mRNA and other messages containing similar AREs, such as granulocyte-macrophage colony stimulating factor (GM-CSF) and interleukin-3 (IL-3), by binding directly to the AREs of these messages and destabilizing them in some way (Carballo et al., 1998).

Regulation of mRNA stability is of critical impor-

tance in early *Xenopus* development, partly because no zygotic transcription occurs until the mid-blastula transition (MBT), several hours and several hundred cell divisions after egg fertilization. Thus, early embryonic development in this species relies on the presence of maternal mRNAs, whose stability is therefore crucial to this process.

For these reasons and because of the relative ease with which specific proteins can be introduced and expressed (Moon and Christian, 1989) or 'knocked out' (Dash et al., 1987; Shuttleworth and Colman, 1988; Heasman et al., 1994; Wylie et al., 1996; Kofron et al., 1997) in early development, we attempted to isolate members of the CCCH class of zinc finger proteins in *Xenopus*. In this report, we describe the cloning of four members of this class in this organism, three of which appear to be homologues of the three previously identified mammalian proteins of the double CCCH zinc finger class. We also isolated a previously unidentified member of this family, which, instead of the usual two  $Cx_8Cx_5Cx_3H$  motifs, contained a total of four such motifs. This transcript appeared to be highly expressed only in the ovary, oocyte, egg, and early embryo. The other three mRNAs were also maternally expressed to varying degrees, suggesting potential roles for all four family members in mRNA stability during oocyte maturation and/or early embryogenesis.

## 2. Materials and methods

### 2.1. Screening the *Xenopus* kidney library

Reverse transcriptase-polymerase chain reaction (RT-PCR) was performed on *Xenopus* egg poly-A<sup>+</sup> RNA using degenerate primers based on conserved regions of mammalian CCCH zinc finger motifs. PCR products were subcloned using the TA cloning kit (Invitrogen, San Diego, CA), then cloned into the EcoR1 site of the pCR2.1 vector (Stratagene, La Jolla, CA). One of the PCR products, X50, was used to screen a *Xenopus* kidney cDNA library (Stratagene). Positive clones were grouped according to the sizes of their restriction fragments, and the similarities of fragments that hybridized to the original PCR probe. Subclones were sequenced by the dideoxy-chain termination method, and sequence alignments were performed using MacVector 6.0 computer software (Oxford Molecular Group, Campbell, CA).

### 2.2. Northern blotting

These were performed as described (Lai et al., 1990), using random-primed <sup>32</sup>P-labeled *Xenopus* cDNA fragments as probes. Roughly equivalent loading was confirmed by acridine orange staining of the gels prior to

transfer. Where indicated, mRNA levels on northern blots were quantified using a Phosphorimager (Molecular Dynamics, Sunnyvale, CA).

### 2.3. *In situ* hybridization

Antisense and sense digoxigenin-labeled RNA probes representing bp 1–245 and 245–565 of XC3H-4 were used for *in situ* hybridization histochemistry as described (Zeldin et al., 1997), except that hybridization and washes were performed at 48°C, and the colorimetric reaction was performed without 10% polyvinyl alcohol.

## 3. Results

### 3.1. Screening the *Xenopus* kidney cDNA library

Screening of a *Xenopus* kidney cDNA library with the 189 bp PCR probe yielded 152 positive plaques. 62 of these were purified and subjected to restriction mapping and Southern analysis. Inserts of identical size yielding identical restriction fragments that hybridized to the probe were grouped together. One clone from each of 15 groups was selected for sequencing, and in each case the conserved YKTELC<sub>8</sub>Cx<sub>5</sub>Cx<sub>3</sub>H motif was identified in the predicted amino acid sequence. The initial 15 cDNA sequences were obtained using universal primers on the complete insert. If different size inserts shared identical sequences, they were also grouped together. Next, the largest cDNA insert from each of the five remaining groups was subcloned into pBluescribe as described in Section 2 and sequenced, initially using universal primers, and subsequently with internal primers. Sequences for complete open reading frames coding for four distinct proteins were obtained by this means, and are described below.

### 3.2. Cloning and characterization of XC3H-1, a TTP homologue

We have adopted a simplified nomenclature for these proteins, with X referring to *Xenopus*, and C3H referring to the CCCH class of zinc fingers. XC3H-1 contained a 2756 bp insert that included a single open reading frame encoding a protein of 313 amino acids with a predicted molecular mass of 34.8 kDa, pI 9.3. This insert contained 19 bp of 5' untranslated region (UTR) and 1998 bp of 3' UTR (GenBank accession number AF061980). The predicted amino acid sequence contained three highly conserved PPPPG-like motifs; this fact, along with an 84% amino acid identity in the 63 residue region containing the two CCCH motifs, and an overall 49% amino acid identity to human TTP (Taylor et al., 1991), suggested that this protein was the *Xenopus* homologue of mammalian TTP (Fig. 1). An alignment of the

XC3H-1 amino acid sequences with sequences of other mammalian (mouse, rat, and bovine) TTP homologues (Lai et al., 1990; Taylor et al., 1995) revealed that XC3H-1 exhibited a 48–50% amino acid identity with each (data not shown). In addition, serine 220 in mouse TTP (Lai et al., 1990), previously identified as a p42 MAP kinase phosphorylation site (Taylor et al., 1995), was also conserved in XC3H-1 as serine 222.

The ~4kb XC3H-1 transcript was widely expressed in adult *Xenopus* tissues (Fig. 2A); its expression was detectable in the oocyte and egg, but not in the early embryo, and its expression did not become prominent until after neural tube formation in embryonic development (Fig. 2B). Its expression in the ovary was considerably greater than that of the oocyte and egg per amount of RNA (Fig. 2A and B), suggesting that XC3H-1 might be highly expressed in the ovarian follicle cells or surrounding interstitium.

### 3.3. Cloning and characterization of XC3H-2, a cMG1 (TIS11B, ERF-1) homologue

XC3H-2 contained a 2896 bp insert (GenBank accession number AF061981) that predicted a single open reading frame encoding a protein of 345 amino acids, with a predicted molecular mass of 38.3 kDa, pI 9.1. This insert contained 240 bp of 5' UTR and 1621 bp of 3' UTR. In the 63 amino acid region containing the two CCCH repeats, the predicted amino acid sequence was 98% identical to that of rat cMG1, murine TIS11B, and human ERF-1 (Gomperts et al., 1990; Varnum et al., 1991; Barnard et al., 1993); the overall amino acid identity between the *Xenopus* and the human protein was 76% (Fig. 3). Northern analysis revealed widespread expression of one transcript of about 2.7 kb in adult *Xenopus* tissues (Fig. 4A). However, during the early stages of embryonic development, XC3H-2 was expressed as multiple transcripts (~3 kb, ~2.5 kb, ~1.3 kb). The adult ~4 kb transcript was expressed only in the later stages of development following the mid-blastula transition (MBT), when zygotic transcription is initiated (Fig. 4B). As discussed in more detail below, the XC3H-4 mRNA was approximately the same size as the ~1.3 kb transcript that hybridized to the XC3H-2 probe present in the egg and in embryonic stages 7 and 9; this fact, along with the high level expression of XC3H-4 in the ovary and early embryo, suggested the possibility that the 1.3 kb species represented cross-hybridization of the XC3H-2 probe to XC3H-4. However, when a 1.6 kb Bam HI fragment of XC3H-2 that contained most of the 3' UTR was used as a probe, it also hybridized to the 1.3 kb transcript (data not shown), indicating that this transcript was one of the multiple transcripts detected in the oocyte, egg and early embryo, and was not the XC3H-4 transcript.



XC3H-1	1	MSSILDIHTLYQNLRLNLDLSEDLDSP-REG	29
hTTP	1	MDLTAIYESL--LSLSPDVPVPSPDHG	24
XC3H-1	30	KLLSTQRRHSCTPELDDLFRPSSDTWNYDL	59
hTTP	25	GTESSPGWGSSGPWS---LSPSDSSSPSGVT	51
XC3H-1	60	LRTPFRRSDRSISLTEGSRLAFPAPPPGFPP	89
hTTP	52	SRLPGR---STSLVEGRSCGWV <u>PPPPGFAP</u>	78
XC3H-1	90	LKTALPALPAPSP-----RYKTEL	108
hTTP	79	LAPRLGPELSPSPSTPTATSTTPSRYKTEL	108
XC3H-1	109	<span style="border: 1px solid black;">C</span> RTFSETGT <span style="border: 1px solid black;">C</span> KYGAK <span style="border: 1px solid black;">C</span> QFA <span style="border: 1px solid black;">H</span> GKIELREP NR	138
hTTP	109	<span style="border: 1px solid black;">C</span> RTFSESGR <span style="border: 1px solid black;">C</span> RYGAK <span style="border: 1px solid black;">C</span> QFA <span style="border: 1px solid black;">H</span> GLGELRQANR	138
XC3H-1	139	HPKYKTEL <span style="border: 1px solid black;">C</span> HKFYLYGEC <span style="border: 1px solid black;">P</span> YGSRC <span style="border: 1px solid black;">N</span> FI <span style="border: 1px solid black;">H</span> HP	168
hTTP	139	HPKYKTEL <span style="border: 1px solid black;">C</span> HKFYLYGRC <span style="border: 1px solid black;">P</span> YGSRC <span style="border: 1px solid black;">H</span> FI <span style="border: 1px solid black;">H</span> NP	168
XC3H-1	169	RE---QGT SQHILRQSLSYSGVPT-RRGSP	194
hTTP	169	SEDLAAPGHPPVLRQSI SFSGLP SGRRTSP	198
XC3H-1	195	PPPGLPDPAAFSR--APSVSPPPSSDLIFS	222
hTTP	199	<u>PPPGLAGPSLSSSSSFSPSS</u> <u>PPPPGDLPLS</u>	228
XC3H-1	223	PIPTEARSHVSSSLRSADSYSHCCSCHCSRA	252
hTTP	229	PSAFSAAPGTPLAR-RDPTPVCCPS-CRRA	256
XC3H-1	253	GTITQDLLSTQMLLRSPSCSSLP-ETECYS	281
hTTP	257	TPIS-VWGPLGGLVTRTPSVQSLGSDPDEYA	285
XC3H-1	282	S-----GSESPVFEQSYQSPPP---SNKR	302
hTTP	286	SSGSSSLGSDSPVFEAGVFAPPQPVAAARR	315
XC3H-1	303	LPIFNRLSVSD	313
hTTP	316	LPIFNRLSVSE	326

Fig. 1. Predicted amino acid sequence of XC3H-1 aligned with human TTP. The human DNA sequence is from Taylor et al. (1991) (Genbank accession number M63625); the *Xenopus* DNA sequence is in Genbank (accession number AF061980). The CCCH residues in the two zinc finger repeats are delineated by boxes; residues that are identical in both proteins are shaded in gray. The three PPPPG groups in the mouse sequence are underlined. Amino acid positions are numbered on both sides. The alignment was performed using the ClustalW Alignment program in MacVector 6.0 computer software (Oxford Molecular Group, Campbell, CA).

### 3.4. Cloning and characterization of XC3H-3, a TIS11D homologue

XC3H-3 contained a 1.6 kb insert that predicted an incomplete open reading frame. The 5' end of this coding sequence was sought by RT-PCR on *Xenopus* egg RNA, using a 5' internal primer derived from previous 5' RACE of X44 as described in Section 2, whose partial sequence was identical to that of XC3H-3 and shared significant sequence similarity to ERF-2 (TIS11D) (Varnum et al., 1991; Nie et al., 1995). The 3' internal primer contained a unique SphI enzyme site and was about 205 bp from the 5' end of XC3H-3. A full-length clone was generated by substituting the 443 bp of PCR

product for the first 205 bp BamHI/SphI fragment of the incomplete clone.

The resulting 1825 bp of DNA sequence (GenBank accession number AF061982) contained an open reading frame encoding a protein of 364 amino acids with a predicted molecular mass of 40.3 kDa, pI 9.3. The 3' UTR consisted of 733 bp and ended with a poly A tail. The predicted amino acid sequence of XC3H-3 resembled that of the third known mammalian member of the family, TIS11D (ERF-2) (Varnum et al., 1991; Nie et al., 1995; Phillips and Blackshear, unpublished data), with 97% amino acid identity over the 63 residue double zinc finger region and an overall 71% amino acid identity with the human homologue, ERF-2 (Fig. 5). XC3H-3

Fig.  
emb  
XC3  
dete

was  
tha  
the  
bet  
(V  
not  
aci  
ori  
sio  
poi  
cor  
mo  
of  
XC  
the  
init  
wa  
et :

in  
anc  
the  
of  
lik  
scr  
rel  
7  
eve  
tra



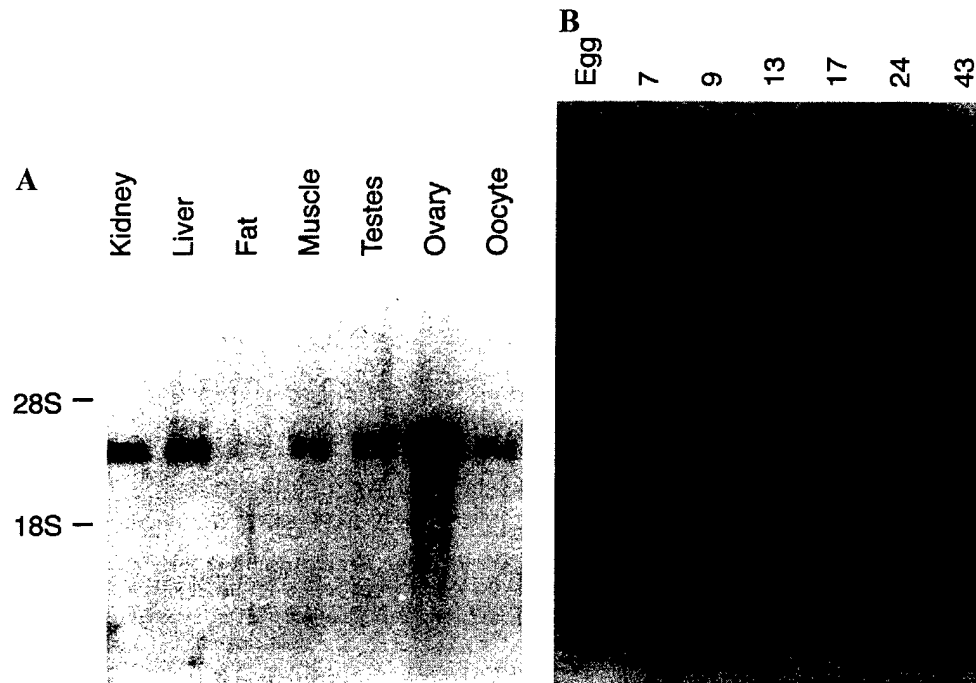


Fig. 2. Expression of XC3H-1 mRNA. Northern analysis was performed on total RNA (15 µg/lane) prepared from (A) adult tissues and (B) embryos at different stages of development. The positions of the major species of ribosomal RNA are indicated. Blots were probed with the 1.6 kb XC3H-1 cDNA. The ~4kb XC3H-1 transcript was widely expressed in the adult. It was detectable in the oocyte and egg, but only became detectable in the embryo at stages 24 (tailbud) and 43 (tadpole).

was more similar to human ERF-2 (Nie et al., 1995) than to mouse TIS11D (Varnum et al., 1991) at both the amino and carboxyl termini. The loss of alignment between the carboxyl terminal sequences of TIS11D (Varnum et al., 1991) and ERF-2 (Nie et al., 1995) was noted previously (Nie et al., 1995) to begin 12 amino acids before the TIS11D stop codon. Inspection of the original TIS11D nucleotide sequence (Genbank accession number M58564) revealed that a frameshift at this point in the TIS11D sequence would have resulted in a continuation of the open reading frame that was much more similar to ERF-2 and XC3H-3. Similarly, the loss of amino terminal homology between TIS11D and XC3H-3 on the one hand and ERF-2 and TIS11D on the other appears to be due to the fact that the earlier initiator methionine used by both ERF-2 and XC3H-3 was not present in the cDNA encoding TIS11D (Varnum et al., 1991).

XC3H-3 was expressed as a major ~5 kb transcript in all of the adult *Xenopus* tissues analyzed (Fig. 6A), and it was readily detectable in the egg and again after the MBT (Fig. 6B). The decreases in size and amount of the transcript in embryonic stages 9 and 13 most likely represent the early degradation of maternal transcripts before zygotic transcription. In addition, the relatively minor ~4 kb transcript in the egg and stage 7 embryo may represent a degradation product; however, the possibility of an additional minor or related transcript cannot be ruled out.

### 3.5. Cloning and characterization of a novel member of the vertebrate CCCH family, XC3H-4

The three *Xenopus* members of the CCCH protein family described thus far contained two zinc finger motifs, each with two YKTELC<sub>8</sub>Cx<sub>5</sub>Cx<sub>3</sub>H motifs in which the H of the first finger was separated from the first C of the second finger by 18 amino acids. These proteins were each highly homologous to one of the three known mammalian CCCH proteins with two zinc fingers. XC3H-4, on the other hand, appeared to represent a previously unknown vertebrate family member, with two additional zinc finger motifs and a very restricted expression pattern. XC3H-4 contained a 1195 bp insert (GenBank accession number AF061983) that predicted a single open reading frame encoding a protein of 276 amino acids with a calculated molecular mass of 30 325 Da, pI 5.9 (Fig. 7A). This insert also contained 38 bp of 5' UTR and 309 bp of 3' UTR. Aside from the zinc finger motifs, this protein did not significantly resemble any known protein in current databases, including the most recent releases of expressed sequence tag databases.

The 3' UTR contained four widely spaced AUUUA instability motifs, which can promote rapid turnover of the mRNA (Shaw and Kamen, 1986). Besides the expected two highly conserved YKTELC<sub>8</sub>Cx<sub>5</sub>Cx<sub>3</sub>H motifs, XC3H-4 also contained two additional CCCH motifs with identical internal spacing but without the

XC3H-2	1	MSTALISPTIFDLSDVLCKSNKMLNYYNNNI	30
ERF1	1	MTTTLVSAITFDLSEVLCKGNKMLNYS---	27
XC3H-2	31	INPSTTNFPLMDRKA VGT PAIVGFPRRHSV	60
ERF1	28	-APSAG-GCLLDRKA VGT PAGGGGFPRRHSV	55
XC3H-2	61	TLPNAKFNQNFQFLNSLKMEPSTAMGNKENK	90
ERF1	56	TLPSSKFHQNQLLSSSLKGEPAPALSSRDSR	85
XC3H-2	91	FRDRSFSESGERLLQK---PGG-QVNSSRY	116
ERF1	86	FRDRSFSEGGERLLPTQKQPGGGQVNSSRY	115
XC3H-2	117	KTEL[CR]PFEEENGSC[CKYGDK]CQFA[HG]IHELRL	146
ERF1	116	KTEL[CR]PFEEENGAC[CKYGDK]CQFA[HG]IHELRL	145
XC3H-2	147	SLTRHPKYKTEL[CR]TFHTIGF[CPYGP]R[CH]F	176
ERF1	146	SLTRHPKYKTEL[CR]TFHTIGF[CPYGP]R[CH]F	175
XC3H-2	177	I[HN]AEERRLVSGRDQAHFSLSSSSSKMERPR	206
ERF1	176	I[HN]AEERRALAG-----ARDLSAD-----RPR	197
XC3H-2	207	LQHSFSFAGFPPTN-----GLLDSPTSIT	230
ERF1	198	LQHSFSFAGFPSPAATAAATGLLDSPTSIT	227
XC3H-2	231	PPPIILSTDDLINSPTLHDCSTNPFTFSSQE	260
ERF1	228	PPPIILSADDLLGSPTLPDGTNNPFAFSSQE	257
XC3H-2	261	LASLFAPSMGMQMPLSNSNASGSPTSFLFR	290
ERF1	258	LASLFAPSMGLPG-----GGSPTTFLFR	280
XC3H-2	291	PMSESPQMFDSPSPRDSLSDQEGYLSSSS	320
ERF1	281	PMSESPHMFDSPPSPQDSLSDQEGYLSSSS	310
XC3H-2	321	S---GSDSPTLDTTKRLPIFSRLSISDD	345
ERF1	311	SSHSGSDSPTLDNSRRLPIFSRLSISDD	338

Fig. 3. Predicted amino acid sequence of XC3H-2 aligned with human ERF-1. The human DNA sequence is from Barnard et al. (1993) (GenBank accession number X71901); the *Xenopus* DNA sequence is in Genbank (accession number AF061981). All other symbols are as in the legend to Fig. 1.

YKTEL lead-in sequence (Fig. 7A). These two fingers were separated by only seven amino acids (Fig. 7A). In the second, more carboxyl-terminal pair of CCCH repeats, the lead-in sequence was PYRER in the first repeat and SARET in the second repeat (Fig. 7B). A glutamic acid residue was conserved in the fourth position in the lead-in sequence from all four putative zinc fingers, and the remaining four residues were either basic or neutral. The consensus shared by the four CCCH repeats in XC3H-4 was therefore ExCx<sub>6</sub>GxCxYx<sub>3</sub>CxFxH (where *x* is a variable amino acid) (Fig. 7B).

Northern analysis revealed high-level expression of the 1.2 kb XC3H-4 transcript in the ovary, but there was no detectable expression in any other of the adult tissues examined (Fig. 8A). No detectable expression was observed in the kidney by this means, suggesting the possibility that the original kidney library from

which these cDNAs were selected might contain contaminating oocyte sequences. XC3H-4 mRNA was very highly expressed in the oocyte, egg, and early embryo until stage 10, corresponding to the end of MBT, but it had disappeared by stage 13 (Fig. 8B). To evaluate the expression of XC3H-4 between stages 9 and 13 more carefully, embryos at stages 9–13 were collected, and RNA from these embryos was analyzed on a northern blot (Fig. 8C). The rapid disappearance of the mRNA after stages 9–10 suggests rapid degradation of maternal mRNA without new zygotic transcription.

### 3.6. Localization of XC3H-4 mRNA in *Xenopus* ovary by *in situ* hybridization

*In situ* hybridization of formalin-fixed, paraffin-embedded *Xenopus* ovary sections with the bp 1–245 digoxigenin-labeled antisense mRNA revealed strong

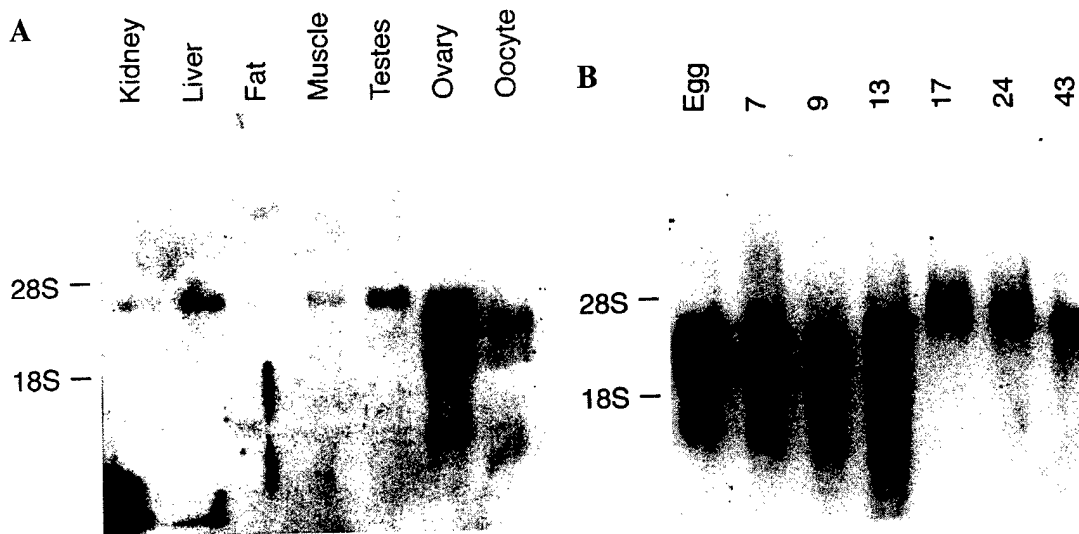


Fig. 4. Expression of XC3H-2 mRNA. Northern analysis was performed on total RNA (15 µg/lane) prepared from (A) adult tissues and (B) embryos at different stages of development. Blots were probed with the 0.5 kb Bam HI XC3H-2 cDNA fragment. Multiple transcripts, ~3 kb, 2.5 kb, and 1.3 kb in size, were present in the ovary, egg and early embryo. The 4 kb transcript that is widely expressed in the adult was first detectable in the embryo at stage 13 (late gastrula).

positive staining in the cytosol of oocytes at all developmental stages examined (Fig. 9). There was no detectable nuclear staining, and no apparent staining of the ovarian stromal tissue. A similar staining pattern was observed with the bp 245–363 probe (not shown). Although the smaller oocytes appeared to be uniformly stained in the cytosol, more mature oocytes exhibited clear animal pole localization (Fig. 9).

#### 4. Discussion

By screening a *Xenopus* kidney cDNA library with degenerate PCR probes based on mammalian sequences, we have cloned a new vertebrate member of the double CCCH class of zinc finger proteins, in addition to the apparent *Xenopus* homologues of TTP, cMG1, and TIS11D. The unusual properties of this novel protein, which we have labeled XC3H-4, include two additional CCCH zinc finger motifs without the usual YKTEL lead-in sequence. In addition, expression of its mRNA appeared to be limited to oocytes, eggs, early embryos (before stage 10) and adult ovary; in all these tissues, its transcripts appeared to be quite abundant. These transcripts were present in oocytes of all stages, and were localized to the animal pole in mature oocytes.

Unlike XC3H-1, XC3H-2, and XC3H-3, which all have basic *pI* values, XC3H-4 is an acidic protein with a *pI* of 5.9; secondary structure analysis suggests that it has no notable hydrophobic domains characteristic of integral membrane proteins. Proline, serine and leucine each comprise 12–13% of the amino acid composition. Apart from its four CCCH motifs, XC3H-4 does not appear to be related to any sequences in currently

available databases; specifically, no mammalian counterparts of this protein were found in current databases. Its function remains unknown; however, it may be possible to explore the function of XC3H-4 during oogenesis and/or early development by the use of antisense oligonucleotide and host transfer techniques (Wylie et al., 1996).

Recently, our group has established a possible function for the subclass of CCCH proteins to which all four of the cloned *Xenopus* proteins belong. TTP, the prototype of proteins of this class, was shown to be necessary for the normal lability of TNFα mRNA in mouse macrophages (Carballo et al., 1998). When TTP is absent, as in TTP-knockout mice, this abnormal stabilization of the TNFα mRNA results in hypersecretion of TNFα from macrophages, and a chronic inflammatory syndrome in intact mice (Taylor et al., 1996a,b; Carballo et al., 1997, 1998). TTP has been shown to bind directly to the AU-rich element in the TNFα mRNA 3' UTR (Carballo et al., 1998); this binding in some way results in destabilization of the mRNA (Carballo et al., 1998), probably by initiating its deadenylation (Lai et al., submitted). These properties of mammalian TTP, i.e. direct binding to the AREs of TNFα and other cytokines and the ability to destabilize TNFα mRNA in co-transfection studies, were shared by another mammalian family member, cMG1, and were also shared by XC3H-1 and XC3H-3 (Lai et al., submitted). There was no obvious response to XC3H-2 in these studies, perhaps because of its relatively poor expression in the 293 cell transfection experiments (Lai et al., submitted). XC3H-4 was also ineffective in these assays, perhaps for the same reason, or perhaps because it differs from the three known mammalian proteins in

XC3H-3	1	MSATLLSAFYDIDLLYKNEKALNNLALSTM	30
ERF2	1	MSTLLSAFYDVDFLCKTEKSLANLNLNM	30
XC3H-3	31	LDKKAVGSPVSSSTNSN-LFPGFLRRHSASN	59
ERF2	31	LDKKAVGTPVAAAPSSSGFAPGFLRRHSASN	60
XC3H-3	60	LQALSGSTNPAKFCHNNNNNQLN-----	82
ERF2	61	LHALAHPAPSPGSCSPKFPGAANGSSCGSA	90
XC3H-3	83	-----ESAASSTALLNRENKFR	99
ERF2	91	AAGGPTSYGTLKEPSGGGGTALLNKENKFR	120
XC3H-3	100	DRSFSENGERSQHLLHLQQQQQQQKAGAHV	129
ERF2	121	DRSFSENGDRSQHLLHLQQQQKGG-GGSQI	149
XC3H-3	130	NSTRYKTEL[CRPFEENGAC]KYGEK[CQFAHG	159
ERF2	150	NSTRYKTEL[CRPFEESGT]KYGEK[CQFAHG	179
XC3H-3	160	FHELRSLTRHPKYKTEL[CRTFHTIGFC]PYG	189
ERF2	180	FHELRSLTRHPKYKTEL[CRTFHTIGFC]PYG	209
XC3H-3	190	PR[CHFIH]NAEERRQAPGAGE-----	209
ERF2	210	PR[CHFIH]NADERRPAPSGGASGDLRAFGTR	239
XC3H-3	210	-----RPKLHHSLSFSGFPNHS LD	228
ERF2	240	DALHLGFPREPRPKLHHSLSFSGFP SGHHQ	269
XC3H-3	229	SP-----LLESPTSRTPPPQ-----	243
ERF2	270	PPGGLESPLLLDSPTSRTPPPPSCSSASSC	299
XC3H-3	244	SSGSLYCQELLQLNNNN-----	260
ERF2	300	SSSASSCSSASAASTPSGTPTCCASAAAAL	329
XC3H-3	261	-----P-----CANN	265
ERF2	330	RLLYGTGGAEDLLAPGAPCAACSSASCANN	359
XC3H-3	266	AFTFSGQELGLIAPLAIHTQN-----QSY	289
ERF2	360	AFAFGPESLITPLAIQTHNFAAVAAAAY	389
XC3H-3	290	CR-----OPCS-----	295
ERF2	390	YRSQQQQQQQGLAPPAQPPAPPSATLPAGA	419
XC3H-3	296	----SPPLSFQPLRRVSESPVFDAPPSPPD	321
ERF2	420	AAPPSPPFQPLRRLSDSPVFDAPPSPPD	449
XC3H-3	322	SLSDRDSYLSGSLSSGSLSGSDSPTLDSNR	351
ERF2	450	SLSDRDSYLSGSLSSGSLSGSESPSLDPGR	479
XC3H-3	352	RLPIFSRLSISDD	364
ERF2	480	RLPIFSRLSISDD	492

Fig. 5. Predicted amino acid sequence of XC3H-3 aligned with human ERF-2. The human DNA sequence is from Nie et al. (1995) (GenBank accession number X78992). The *Xenopus* DNA sequence is in Genbank (accession number AF061982). All other symbols are as in the legend to Fig. 1.

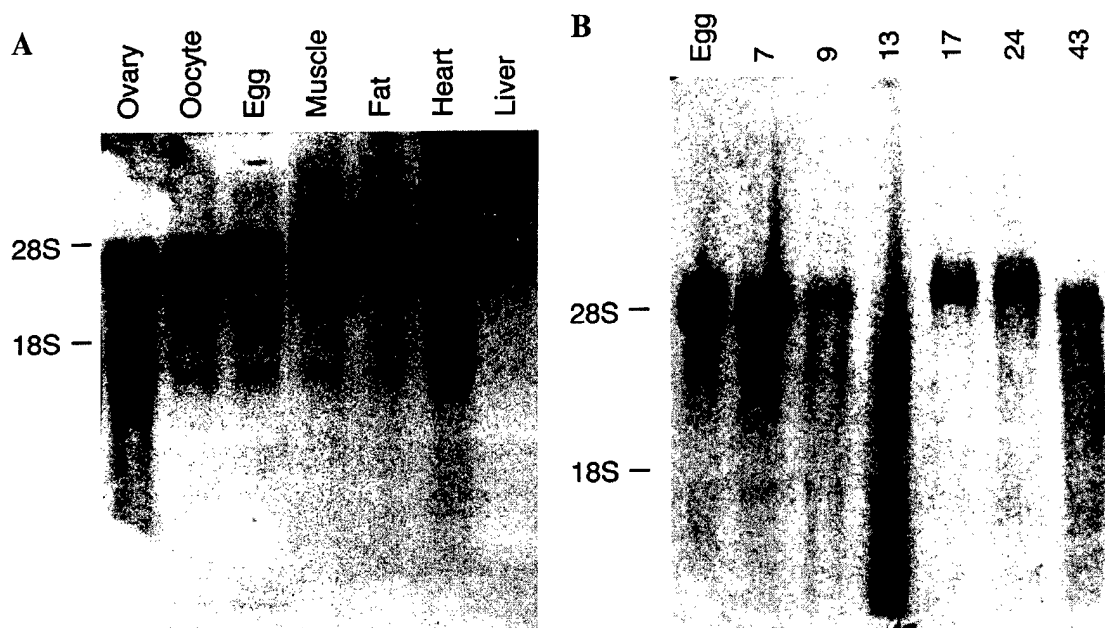


Fig. 6. Expression of XC3H-3 mRNA. Northern analysis was performed on total RNA (15 µg/lane) prepared from (A) adult tissues, oocytes, and eggs and (B) embryos at different stages of development. Blots were probed with a (A) 0.5 kb EcoRI/NcoI XC3H-3 cDNA and (B) 0.8 kb SstI–HindIII XC3H-3 cDNA. The ~4.4 kb transcript was expressed in all adult tissues tested, as well as oocytes and eggs; it was also expressed in the embryo by stage 19.

several ways, as noted above. However, it seems likely that at least three, and possibly all four, CCCH proteins from *Xenopus* will turn out to be ARE-binding and mRNA destabilizing proteins. Since control of poly A tail length is critical in early development in *Xenopus* and other species (Beelman and Parker, 1995; Audic et al., 1997; Paillard et al., 1998), it seems likely that one of the functions of this class of proteins in *Xenopus* is to regulate mRNA turnover at critical times during oocyte maturation or early embryo development. The specific mRNA targets of each of these proteins will be the subject of further study.

In situ hybridization of *Xenopus* ovary using digoxigenin-labeled antisense RNA probes for XC3H-4

revealed that the smallest oocytes were strikingly and uniformly labeled in the cytosol but not the nucleus. More mature oocytes showed strong positive cytoplasmic staining in the animal pole, including the cortex and the perinuclear region. Only a small fraction of *Xenopus* maternal RNAs are thought to be localized in this way (Rebagliati et al., 1985). The mechanisms of this localization of XC3H-4 to the animal cortex, and its developmental significance, will require further study.

To date, the three known vertebrate double C<sub>x</sub>C<sub>x</sub>C<sub>x</sub>H proteins have contained two very similar CCCH zinc fingers, with the carboxyl-terminal H of the first finger separated from the amino terminal C of the second finger by 18 amino acids in all three cases. This

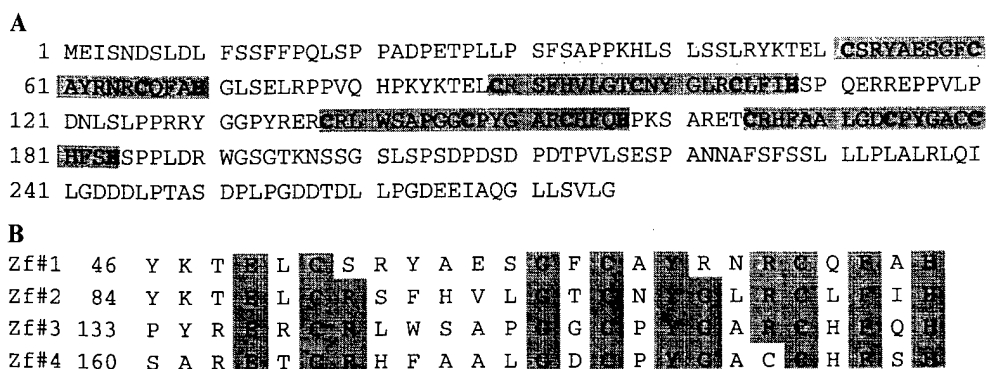


Fig. 7. Predicted amino acid sequence of XC3H-4. The DNA sequence has been deposited in GenBank (accession number AF061983). (A) The predicted amino acid sequence encoded by XC3H-4 cDNA is shown here. The four putative zinc finger motifs are shaded in gray; the CCCH residues of the four zinc finger repeats are delineated by bold letters. Amino acid positions are numbered on the left. (B) Alignment of the four zinc finger motifs in XC3H-4, from the most amino terminal (1) to the most carboxy terminal (4). All other symbols are as in the legend to Fig. 1.

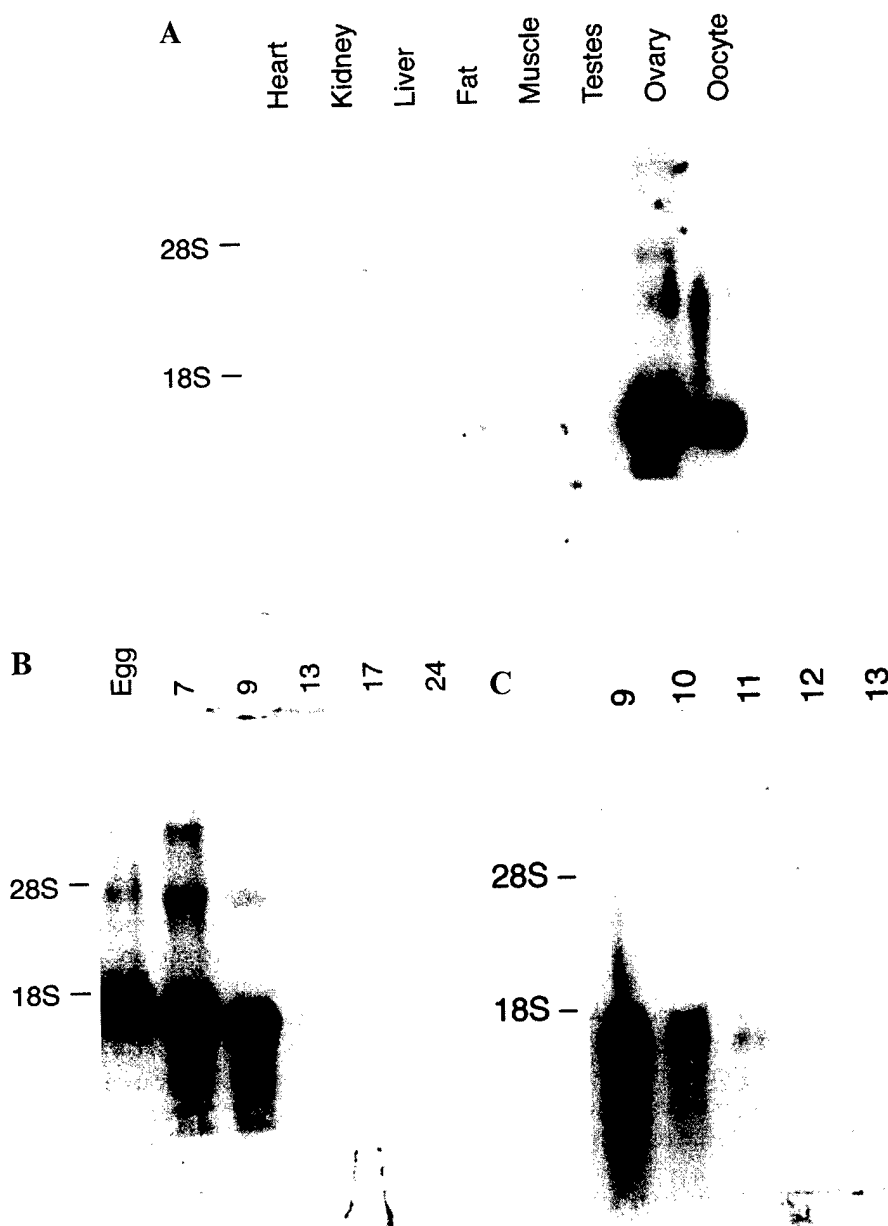


Fig. 8. Expression of XC3H-4 mRNA. Northern analysis was performed on total RNA prepared from (A) adult tissues (15  $\mu$ g/lane), (B) embryos at different stages in development (15  $\mu$ g/lane), and (C) embryos from stages 9–13 (7  $\mu$ g/lane). Blots were probed with the 0.8 kb PstI XC3H-4 cDNA. Expression of the  $\sim$ 1.2 kb transcript was limited to the ovary, egg and early embryo until stage 10 (late blastula); there was no detectable expression after the mid-blastula transition.

spacing was exactly the same in the *Xenopus* homologues; in fact, 14 of the 18 intervening amino acids from XC3H-1 were identical to those of human TTP (Taylor et al., 1991), whereas all 18 of the residues were identical when comparing XC3H-2 with ERF-1 (Barnard et al., 1993) and XC3H-3 to ERF-2 (Nie et al., 1995). In contrast, although the amino-terminal two fingers of XC3H-4 were separated by the usual 18 amino acids, the carboxyl-terminal pair of putative zinc fingers was separated by only seven amino acids. This difference between the two pairs of zinc fingers, the different lead-in sequences in the carboxyl terminal pair,

and the overall *pI*, suggest that XC3H-4 is the prototype of a new subclass of CCCH zinc finger proteins.

Recently, a CCCH bovine protein containing four zinc fingers in the  $Cx_7Cx_5Cx_3H$  configuration and one zinc finger in the  $Cx_8Cx_4Cx_3H$  form, in addition to a  $Cx_2Cx_4Hx_4C$  zinc knuckle motif, was identified as the 30 kDa subunit of cleavage and polyadenylation specificity factor (CPSF) (Barabino et al., 1997). None of the five putative zinc fingers had classical TTP  $Cx_8Cx_5Cx_3H$  spacing, and none had the typical YKTEL lead-in sequence (Barabino et al., 1997). CPSF 30K was shown to be essential for 3' end processing of pre-

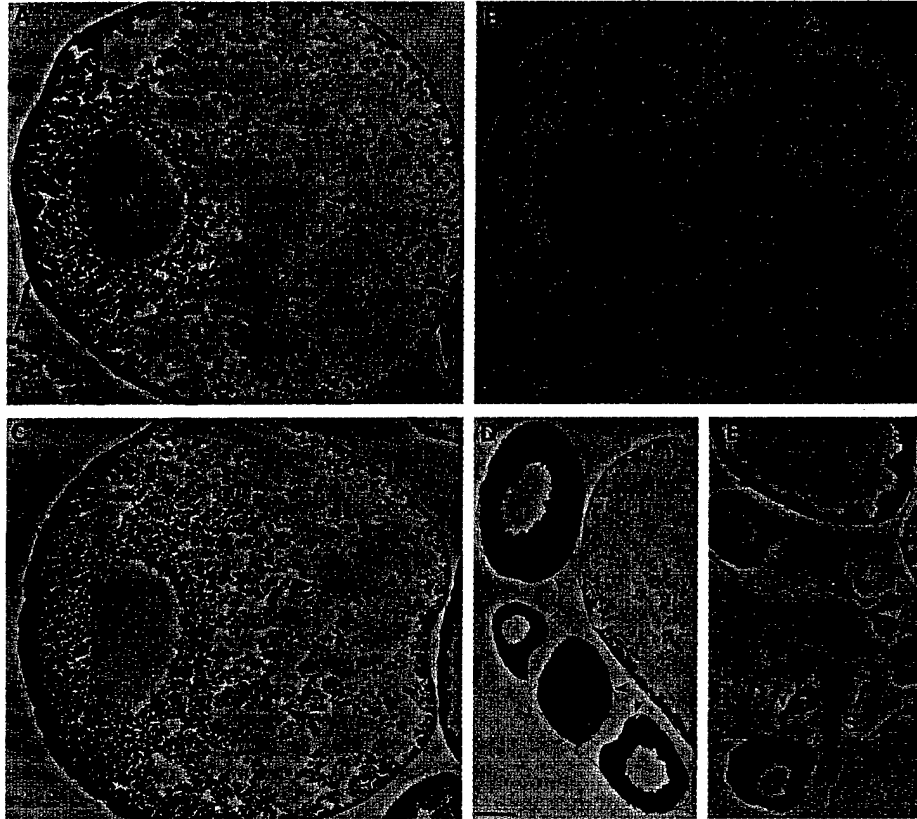


Fig. 9. In situ hybridization histochemistry of XC3H-4 mRNA expression in *Xenopus* ovary. In situ hybridization analysis was performed using antisense (A, C, D, and E) and sense (B) digoxigenin-labeled XC3H-4 mRNA probes on formalin-fixed, paraffin-embedded *Xenopus* ovary sections. Blue color represents the hybridization signal. A and B are neighboring sections of the same oocyte. A, animal pole; V, vegetal pole; N, nucleus; O, ovary stromal tissue. The arrowheads in C–E point to developing oocytes. All photomicrographs were taken at the same magnification; the bar = 100  $\mu$ m.

mRNAs. Although deletion of the zinc knuckle motif resulted in dramatic decreases in the protein's ability to interact with RNA, low affinity for poly (U) was still maintained. These findings support a possible role for the CCCH zinc fingers in binding RNA. The mouse homologue of the 30 kDa subunit of CPSF (GenBank accession number U96448) contained similar CCCH domains but lacked the zinc knuckle.

Based on its 49% amino acid identity and expression pattern, XC3H-1 appears to be the *Xenopus* homologue of mammalian TTP. Like mammalian TTP, XC3H-1 is a basic protein with a *pI* of 9.3. One of the three PPPPG repeats that are present in all mammalian TTPs is perfectly conserved in XC3H-1, while the remaining two repeats are partially conserved. In addition to these conserved repeats, the serine residue that is a major site of MAP kinase phosphorylation, serine 220 in the mouse TTP sequence (Taylor et al., 1995), is also conserved in XC3H-1, as serine 222, as is the following proline. The apparent lack of XC3H-1 expression in early *Xenopus* development makes it feasible to perform overexpression studies in early embryos as an approach to function. Although the mammalian protein has been implicated in the regulation of TNF $\alpha$  release from macrophages

(Taylor et al., 1996a,b; Carballo et al., 1997, 1998), and both TTP and XC3H-1 bind to mammalian TNF $\alpha$  and GM-CSF AREs and destabilize TNF $\alpha$  mRNA (Lai et al., submitted), the physiologically relevant binding partner for XC3H-1 in *Xenopus* remains unknown.

XC3H-2, the *Xenopus* homologue of ERF-1, is unusual in that transcripts of multiple sizes are present during early development. This finding has some precedent in the literature. For example, multiple *DTIS11* transcripts were expressed during *Drosophila* embryogenesis, with a predominant 3 kb species expressed during early embryonic development, and a 6 kb species that was detected in later embryonic stages (Ma et al., 1994). Similarly, alternative splicing in the zinc finger region of the *wil* gene resulted in the major transcript having three extra amino acids between the third and fourth zinc fingers. The major and minor forms then had distinct DNA binding specificities (Bickmore et al., 1992). Thus, certain zinc finger proteins are subject to post-transcriptional regulation that results in the synthesis of multiple potential regulatory proteins from a single transcriptional unit. Future studies will attempt to determine whether the observed multiple transcripts of XC3H-2 are derived from developmentally controlled

embryos  
XC3H-4  
ectable

tototype

g four  
d one  
1 to a  
as the  
speci-  
one of  
TTP  
KTEL  
K was  
f pre-

differential splicing, alternative transcription start sites, or a shortening of maternal mRNA.

The study of zinc finger proteins in *Xenopus* has been focused thus far largely on the C2H2 proteins, which represent a large multigene family whose members are maternally transcribed and widely expressed in eggs and embryos (Koster et al., 1988; Knochel et al., 1989). Oligonucleotide-directed destruction of the 'entire pool' of maternal mRNAs encoding C2H2 zinc finger proteins resulted in no notable change in normal oocyte maturation and embryogenesis (el-Baradi et al., 1991). However, the authors admit to the possibility of incomplete mRNA destruction in their experiments. Our studies have identified four zinc finger proteins of the CCCH class in *Xenopus* that are all represented in the maternal pool of mRNA. Oligonucleotide-mediated selective destruction of transcripts encoding each of these proteins may help to elucidate their potential role in oocyte maturation and/or embryogenesis in *Xenopus*, and perhaps suggest similar roles in early development for their mammalian counterparts.

### Acknowledgements

We are very grateful to Greg Solomon for his help in sequencing cDNA inserts. Supported in part by grants to the US Army Breast Cancer Program and the NIH (PO1 HD17461), both to T.K.B. Jitakshi De was a Howard Hughes Medical Institute Medical Student Research Training Fellow during the course of this work.

### References

- Audic, Y., Omilli, F., Osborne, H.B., 1997. Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage. *Mol. Cell. Biol.* 17, 209–218.
- Barabino, S.M.L., Hubner, W., Jenny, A., Minvielle-Sebastia, L., Keller, W., 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev.* 11, 1703–1716.
- Barnard, R.C., Pascall, J.C., Brown, K.D., McKay, I.A., Williams, N.S., Bustin, S.A., 1993. Coding sequence of ERF-1, the human homologue of Tis11b/cMG1, members of the Tis11 family of early response genes. *Nucleic Acids Res.* 21, 3580.
- Beelman, C.A., Parker, R., 1995. Degradation of mRNA in eukaryotes. *Cell* 81, 179–183.
- Bickmore, W.A., Oghene, K., Little, M.H., Seawright, A., van Heyningen, V., Hastie, N.D., 1992. Modulation of DNA binding specificity by alternative splicing of the Wilms tumor *wil* gene transcript. *Science* 257, 235–237.
- Carballo, E., Gilkeson, G.S., Blackshear, P.J., 1997. Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (–/–) mice. *J. Clin. Invest.* 100, 986–995.
- Carballo, E., Lai, W.S., Blackshear, P.J., 1998. Feedback inhibition of macrophage tumor necrosis factor- $\alpha$  production by tristetraprolin. *Science* 281, 1001–1005.
- Dash, P., Lotan, I., Knapp, M., Kandel, E.R., Goelet, P., 1987. Selective elimination of mRNAs in vivo: complementary oligodeoxynucleotides promote RNA degradation by an RNase H-like activity. *Proc. Natl. Acad. Sci. USA* 84, 7896–7900.
- DuBois, R.N., McLane, M.W., Ryder, K., Lau, L.F., Nathans, D., 1990. A growth factor-inducible nuclear protein with a novel Cys/His repetitive sequence. *J. Biol. Chem.* 265, 19185–19191.
- el-Baradi, T., Bouwmeester, T., Giltay, R., Pieler, T., 1991. The maternal store of zinc finger protein encoding mRNAs in fully grown *Xenopus* oocytes is not required for early embryogenesis. *EMBO J.* 10, 1407–1413.
- Gomperts, M., Pascall, J.C., Brown, K.D., 1990. The nucleotide sequence of a cDNA encoding an EGF-inducible gene indicates the existence of a new family of mitogen-induced genes. *Oncogene* 5, 1081–1083.
- Guedes, S., Priess, J.R., 1997. The *C. elegans* MEX-1 protein is present in germline blastomeres and is a P granule component. *Development* 124, 731–739.
- Heasman, J., Crawford, A., Goldstone, K., Garner-Hamrick, P., Gumbiner, B., McCrea, P., Kintner, C., Noro, C.Y., Wylie, C., 1994. Overexpression of cadherins and underexpression of  $\beta$ -catenin inhibit dorsal mesoderm induction in early *Xenopus* embryos. *Cell* 79, 791–803.
- Heximer, S.P., Forsdyke, D.R., 1993. A human putative lymphocyte  $G_0/G_1$  switch gene homologous to a rodent gene encoding a zinc-binding potential transcription factor. *DNA Cell Biol.* 12, 73–88.
- Knochel, W., Poting, A., Koster, M., el-Baradi, T., Niefeld, W., Bouwmeester, T., Pieler, T., 1989. Evolutionary conserved modules associated with zinc fingers in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* 86, 6097–6100.
- Kofron, M., Spagnuolo, A., Klymkowsky, M., Wylie, C., Heasman, J., 1997. The roles of maternal  $\alpha$ -catenin and plakoglobin in the early *Xenopus* embryo. *Development* 124, 1553–1560.
- Koster, M., Pieler, T., Poting, A., Knochel, W., 1988. The finger motif defines a multigene family represented in the maternal mRNA of *Xenopus laevis* oocytes. *EMBO J.* 7, 1735–1741.
- Lai, W.S., Stumpo, D.J., Blackshear, P.J., 1990. Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein. *J. Biol. Chem.* 265, 16556–16563.
- Lai, W.S., Carballo, E., Strum, J.R., De, J., Tuttle, J.S., Kennington, E.A., Phillips, R.S., Blackshear, P.J., submitted for publication.
- Ma, Q., Herschman, H.R., 1991. A corrected sequence for the predicted protein from the mitogen-inducible TIS11 primary response gene. *Oncogene* 6, 1277–1278.
- Ma, Q., Wadleigh, D., Chi, T., Herschman, H.R., 1994. The *Drosophila* TIS11 homologue encodes a developmentally controlled gene. *Oncogene* 9, 3329–3334.
- Mello, C.C., Draper, B.W., Krause, M., Weintraub, H., Priess, J.R., 1992. The *pie-1* and *mex-1* genes and maternal control of blastomere identity in early *C. elegans* embryos. *Cell* 70, 163–176.
- Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., Priess, J.R., 1996. The PIE-1 protein and germline specification in *C. elegans* embryos. *Nature* 382, 710–712.
- Moon, R.T., Christian, J.L., 1989. Microinjection and expression of synthetic mRNAs in *Xenopus* embryos. *Technique* 1, 76–89.
- Nie, X.F., Maclean, K.N., Kumar, V., McKay, I.A., Bustin, S.A., 1995. ERF-2, the human homologue of the murine Tis11d early response gene. *Gene* 152, 285–286.
- Ning, Z.Q., Norton, J.D., Li, J., Murphy, J.J., 1996. Distinct mechanisms for rescue from apoptosis in Ramos human B cells by signaling through CD40 and interleukin-4 receptor: role for inhibition of an early response gene, *Berg36*. *Eur. J. Immunol.* 26, 2356–2363.
- Paillard, L., Omilli, F., Legagneux, V., Bassez, T., Maniey, D., Osborne, H.B., 1998. EDEN and EDEN-BP, a *cis* element and an associated factor that mediate sequence-specific mRNA deadenylation in *Xenopus* embryos. *EMBO J.* 17, 278–287.
- Phillips, R.S., Blackshear, P.J., unpublished data.
- Rebagliati, M.R., Weeks, D.L., Harvey, R.P., Melton, D.A., 1985.



- Identification and cloning of localized maternal RNAs from *Xenopus* eggs. *Cell* 42, 769–777.
- Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Preiss, J.R., Fire, A., 1996. Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature* 382, 713–716.
- Shaw, G., Kamen, R., 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 46, 659–667.
- Shuttleworth, J., Colman, A., 1988. Antisense oligonucleotide-directed cleavage of mRNA in *Xenopus* oocytes and eggs. *EMBO J.* 7, 427–434.
- Taylor, G.A., Lai, W.S., Oakey, R.J., Seldin, M.F., Shows, T.B., Eddy, R.J., Blackshear, P.J., 1991. The human TTP protein: sequence, alignment with related proteins, and chromosomal localization of the mouse and human genes. *Nucleic Acids Res.* 19, 3454.
- Taylor, G.A., Thompson, M.J., Lai, W.S., Blackshear, P.J., 1995. Phosphorylation of tristetraprolin, a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase in vitro. *J. Biol. Chem.* 270, 13341–13347.
- Taylor, G.A., Carballo, E., Lee, D.M., Lai, W.S., Thompson, M.J., Patel, D.D., Schenkman, D.I., Gilkeson, G.S., Broxmeyer, H.E., Haynes, B.F., Blackshear, P.J., 1996a. A pathogenetic role for TNF $\alpha$  in the syndrome of cachexia, arthritis and autoimmunity resulting from tristetraprolin (TTP) deficiency. *Immunity* 4, 445–454.
- Taylor, G.A., Thompson, M.J., Lai, W.S., Blackshear, P.J., 1996b. Mitogens stimulate the rapid nuclear to cytoplasmic translocation of tristetraprolin, a potential zinc finger transcription factor. *Mol. Endocrinol.* 10, 140–146.
- Thompson, M.J., Lai, W.S., Taylor, G.A., Blackshear, P.J., 1996. Cloning and characterization of two yeast genes encoding members of the CCCH class of zinc finger proteins: zinc finger-mediated impairment of cell growth. *Gene* 194, 225–233.
- Varnum, B.C., Lim, R.W., Sukhatme, V.P., Herschman, H.R., 1989. Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 cells by the tumor promoter tetradecanoyl phorbol acetate. *Oncogene* 4, 119–120.
- Varnum, B.C., Ma, Q., Chi, T., Fletcher, B., Herschman, H.R., 1991. The TIS11 primary response gene is a member of a gene family that encodes proteins with a highly conserved sequence containing an unusual Cys–His repeat. *Mol. Cell Biol.* 11, 1754–1758.
- Warbrick, E., Glover, D., 1994. A *Drosophila melanogaster* homolog of the TIS11 family of immediate early genes that can rescue a *cdrl cdc25* mutant strain of fission yeast. *Gene* 151, 243–246.
- Worthington, M.T., Amann, B.T., Nathans, D., Berg, J.M., 1996. Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proc. Natl. Acad. Sci. USA* 93, 13754–13759.
- Wylie, C., Kofron, M., Payne, C., Anderson, R., Hosobuchi, M., Joseph, E., Heasman, J., 1996. Maternal  $\beta$ -catenin establishes a dorsal signal in early *Xenopus* embryos. *Development* 122, 2987–2996.
- Zeldin, D.C., Foley, J., Goldsworthy, S.M., Cook, M.E., Ma, J., Moomaw, C.R., Tomer, K.B., Steenbergen, C., Wu, S., 1997. CYP2J subfamily cytochrome P450s in the gastrointestinal tract: expression, localization and potential functional significance. *Mol. Pharmacol.* 51, 931–943.

oxy-  
ivity.  
D.,  
Cys/  
ater-  
rown  
MBO  
otide  
icates  
ogene  
esent  
elop-  
Gum-  
1994.  
tenin  
Cell  
ocyte  
zinc-  
-88.  
W.,  
dules  
Acad.  
man,  
n the  
motif  
A of  
stim-  
tein.  
gton,  
on.  
pre-  
ponse  
phila  
gene.  
J.R.,  
lasto-  
reiss,  
eleg-  
on of  
S.A.,  
early  
echa-  
ignal-  
bition  
2363.  
D.,  
nd an  
nyla-

# Establishment of Distinct MyoD, E2A, and Twist DNA Binding Specificities by Different Basic Region-DNA Conformations

THIPHAPHONE KOPHENGNAVONG, JENNIFER E. MICHNOWICZ, AND T. KEITH BLACKWELL\*

Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Received 23 July 1999/Returned for modification 1 September 1999/Accepted 23 September 1999

Basic helix-loop-helix (bHLH) proteins perform a wide variety of biological functions. Most bHLH proteins recognize the consensus DNA sequence CAN NTG but acquire further functional specificity by preferring distinct internal and flanking bases. In addition, induction of myogenesis by MyoD-related bHLH proteins depends on myogenic basic region (BR) and BR-HLH junction residues that are not essential for binding to a muscle-specific site, implying that their BRs may be involved in other critical interactions. We have investigated whether the myogenic residues influence DNA sequence recognition and how MyoD, Twist, and their E2A partner proteins prefer distinct CAN NTG sites (the E-box consensus sequence is underlined). In MyoD, the myogenic BR residues establish specificity for particular CAN NTG sites indirectly, by influencing the conformation through which the BR helix binds DNA. An analysis of DNA binding by BR and junction mutants suggests that an appropriate BR-DNA conformation is necessary but not sufficient for myogenesis, supporting the model that additional interactions with this region are important. The sequence specificities of E2A and Twist proteins require the corresponding BR residues. In addition, mechanisms that position the BR allow E2A to prefer distinct half-sites as a heterodimer with MyoD or Twist, indicating that the E2A BR can be directed toward different targets by dimerization with different partners. Our findings indicate that E2A and its partner bHLH proteins bind to CAN NTG sites by adopting particular preferred BR-DNA conformations, from which they derive differences in sequence recognition that can be important for functional specificity.

A large family of transcriptional regulators is defined by the basic helix-loop-helix (bHLH) motif (40), in which a DNA-binding basic region (BR) lies immediately amino terminal to the HLH dimerization segment (17, 41, 55). In metazoans, bHLH proteins are involved in specification of multiple cell types (33, 43, 56). Some bHLH family members function as homodimers, but others appear to act together with a heterodimeric partner (56). For example, the closely related bHLH proteins that mediate myogenic differentiation, including MyoD, are thought to function as heterodimers with E proteins, a widely expressed bHLH protein subgroup that is exemplified by the E2A proteins (14, 17, 32, 42). Most bHLH protein dimers bind to the consensus CAN NTG (the E box; for ease of identification, the consensus sequence is underlined throughout the text) with each respective BR binding to a half site (9, 19-21, 35, 45, 48). Given the many regulatory processes in which bHLH proteins are involved, the apparent simplicity of the CAN NTG consensus raises the important question of how different bHLH proteins act only on appropriate target genes (56).

In part, the specificity with which bHLH proteins function derives from preferential recognition of different classes of CAN NTG sites by different bHLH protein subgroups. The HLH segment consists of a parallel, left-handed, four-helix bundle (Fig. 1) (19-21, 35, 45, 48). The BR is unstructured in solution (2) but when bound to DNA, it extends N terminally from the HLH segment as an  $\alpha$  helix that crosses the major groove (Fig. 1). Crystallographic analyses have revealed some differences in how these proteins bind DNA. For example, in Myc family and related bHLH proteins, an arginine (Arg) residue at BR position 13 (Fig. 2) specifies recognition of

CACGTG sites (7, 16, 25, 54) by contacting bases in the center (20, 21, 48). However, it still is not understood how bHLH proteins which have a different amino acid at BR position 13 (Fig. 2) bind preferentially to distinct CAN NTG sites (9, 16) or how bHLH proteins establish differences in flanking sequence selectivity (9, 23, 24) that can be of biological importance (1, 30).

Many bHLH proteins that lack R<sub>13</sub>, including MyoD and other E2A partners (Fig. 2), can bind to similar DNA sequences in vitro but act on different tissue-specific genes (56). Cooperative or inhibitory relationships with other transcriptional regulators might contribute to this specificity (34, 39, 46, 58), but it is not likely to derive entirely from other lineage-specific factors, because MyoD can induce myogenesis in many different cell types (56). Initiation of myogenesis by MyoD and other myogenic bHLH proteins depends on three residues that are located within the BR and the BR-HLH junction (A<sub>5</sub>, T<sub>6</sub>, and K<sub>15</sub> [Fig. 1 and 2]). These myogenic residues are not essential for binding a muscle-specific site in vitro or in vivo, suggesting that they are involved in other critical interactions (11, 17, 18, 47, 57). These interactions have been proposed to involve distinct cofactors (11, 17, 57) and the unmasking of an activation domain in MyoD or the myogenic cofactor MEF2 (3, 5, 29, 57). In the MyoD-DNA structure, K<sub>15</sub> is oriented away from the DNA, but A<sub>5</sub> and T<sub>6</sub> face the major groove and could not contact other proteins directly (35) (Fig. 1). However, the latter two residues could influence protein-protein interactions indirectly, by affecting how the BR helix is positioned on the DNA (35). Although substitutions at these positions might not substantially impair binding to particular CAN NTG sites, it is important to determine whether they might have more subtle influences on sequence specificity that could reflect conformational effects.

We have determined that the myogenic residues A<sub>5</sub> and T<sub>6</sub> establish the characteristic MyoD sequence preference, which includes a CAGCTG core. Individual substitutions at these BR

\* Corresponding author. Mailing address: Center for Blood Research and Department of Pathology, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 278-3150. Fax: (617) 278-3131. E-mail: blackwell@cbr.med.harvard.edu.

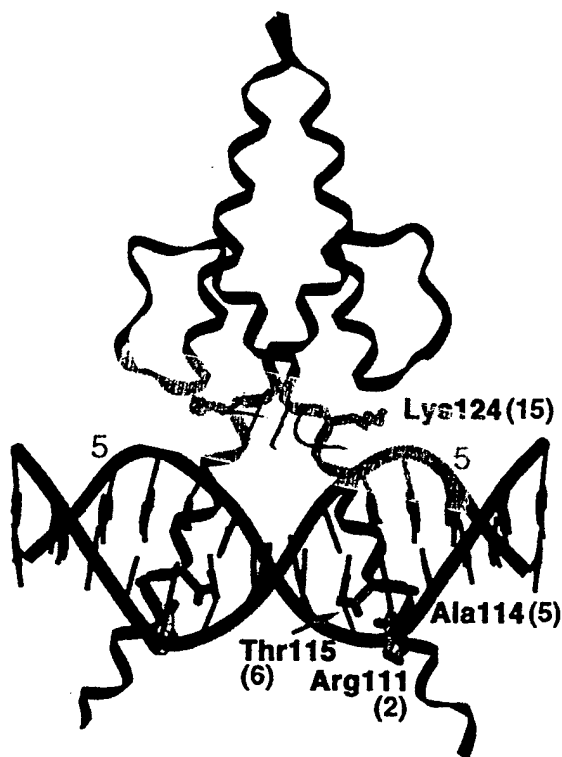


FIG. 1. A MyoD-DNA complex. In this X-ray crystallographic structure (35), a MyoD homodimer is bound to the sequence AAGAGCTGTT, which corresponds to its preferred recognition consensus (9). Residues are numbered as in full-length MyoD, and their positions as specified in Fig. 2 and the text are indicated in parentheses. Binding site positions  $\pm 5$  (numbered as in Fig. 3A) are indicated by grey numerals. Side chains are shown only for the myogenic residues (green) (18) and Arg 111 ( $R_2$ ) (gold).

positions simultaneously alter preferences for multiple bases that MyoD does not contact directly (35), indicating that these preferences are determined indirectly, by how the BR helix is positioned on the DNA. This mechanism is distinct from the standard model for sequence specificity, in which preferred bases are contacted directly (44, 50). The corresponding BR residues are also required for the sequence preferences of E2A proteins, which can recognize either of two distinct half-sites depending on their dimerization partner. E2A homodimers and E2A-MyoD heterodimers bind to asymmetric sites that include a CACCTG core. In contrast, as a heterodimer with the bHLH protein Twist, E2A binds preferentially to half of the symmetric sequence CATATG. The preference of E2A for the former asymmetric sites depends not only on the BR sequence but also on BR positioning that involves the junction region. An analysis of DNA binding by MyoD and E2A junction and BR mutants indicates that a MyoD-like sequence specificity is associated with, but not sufficient for, myogenesis. This supports the model that the BR junction region is also involved in other critical interactions. The results suggest that E2A and its partner bHLH proteins bind DNA by adopting a limited number of preferred BR conformations, each of which is associated with a characteristic DNA sequence preference. They also indicate that binding of cofactors to the MyoD BR can be influenced by how it is positioned on the DNA and are consistent with the idea that relatively subtle differences in binding sequence recognition can modulate bHLH protein activity in vivo.

## MATERIALS AND METHODS

**Mutagenesis, protein expression, and DNA binding assays.** The various MyoD and E2A mutants used in this study have been described previously (17, 18, 57), with the exception of the MyoD mutants shown in Fig. 8A. For construction of those mutants, a *Sall* site that did not alter the encoded amino acid sequence was created at MyoD BR positions 10 and 11 (Fig. 2) by PCR. BR mutants were then generated by PCR using *Pfu* or *Vent* polymerase and introduced into this MyoD (*Sall*) construct as *PmlI-Sall* fragments. Junction region mutations were created similarly by PCR and inserted into MyoD (*Sall*) as *Sall-NarI* fragments. Constructs with both BR and junction mutations were produced by introduction of a mutant *PmlI-Sall* or *Sall-NarI* fragment into the appropriate BR or junction mutant construct. All of these mutations were confirmed by DNA sequencing.

For the in vitro selection experiment shown in Fig. 3, full-length MyoD was expressed in bacteria from a pRK171a-based construct (pT7-MyoD) described previously (53). The MD(E12B), MD(E12B-A), and MD(E12B-AT) mutations (57) were each introduced into this construct within a *PmlI-MluI* fragment. These proteins were expressed by isopropyl- $\beta$ -D-thiogalactopyranoside induction in *Escherichia coli* BL21(DE3)/pLysS cells as described previously (51) and then purified to >90% homogeneity by precipitation in 0.6 M  $(\text{NH}_4)_2\text{SO}_4$ . Precipitated protein was resuspended in a mixture containing 10% glycerol, 20 mM HEPES (pH 7.6), 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1  $\mu$ g of leupeptin per ml, and 1  $\mu$ g of pepstatin per ml.

Other proteins were expressed by in vitro translation (Promega), with in vitro transcription and translation performed in separate steps. Protein expression was carefully quantitated by  $^{35}\text{S}$ -labeled translation and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. These procedures and those for electrophoretic mobility shift assay (EMSA) have been described previously (31). Each EMSA was performed at room temperature and analyzed by autoradiography or phosphorimaging. Individual oligonucleotide sites were 21 bp in length and differed from the MyoD consensus oligonucleotide (7) only at the positions indicated. The MCK-R site corresponds to the right E-box site in the muscle creatine kinase enhancer (18).

**In vitro selection experiments.** Populations of preferred binding sites were isolated by sequential in vitro selection and PCR amplification essentially as described previously (6, 7, 9). During each selection round, DNA that was bound by the protein complex of interest was isolated by EMSA and then amplified by PCR for the next round. In each EMSA selection, care was taken to ensure that sufficient quantities of labeled bound DNA were recovered to maintain a representative population of sequences. These experiments were initiated with 0.5 ng of  $^{32}\text{P}$ -end-labeled starting library. In each subsequent selection round, selections were performed with approximately 0.1 ng of amplified  $^{32}\text{P}$ -body-labeled DNA. In selections for binding to partially purified bacterially expressed MyoD mutant proteins (Fig. 3), these protein preparations were not quantitated, but instead sequential dilutions of these samples were tested for binding. Bound sequences were then recovered and amplified from a sample in which less than 10% of the input DNA was in the bound fraction. This strategy ensured selection of optimal binding sequences. The final selected binding site pool was sequenced directly, using a  $^{32}\text{P}$ -end-labeled primer as described previously (7). Mouse proteins were used in these selections with the exception of Twist, which was from *Xenopus*. Binding site competition analyses (not shown) demonstrated that its binding preferences were indistinguishable from those of mouse Twist, which was used in the EMSA analyses shown.

## RESULTS

**Myogenic BR residues and MyoD DNA binding preferences.** Identification of the myogenic BR residues stemmed originally from studies in which the MyoD BR was replaced with that of E12, a product of the alternatively spliced E2A gene (40). This MyoD mutant [MD(E12B) (Fig. 2)] binds to a muscle-specific regulatory site as a heterodimer with E2A proteins either in vitro or in vivo, but it cannot induce myogenesis in a cell culture assay or activate transcription through a muscle-specific enhancer (17, 57). Resubstitution of the myogenic residues  $A_5$  and  $T_6$  (Fig. 2) in MD(E12B) restores its activity in these functional assays (57). Similar results are obtained when  $A_5$  and  $T_6$  are mutated within MyoD (18, 29, 57) and when analogous substitutions are made in the context of the myogenic bHLH protein myogenin (11). These experiments implicate  $A_5$  and  $T_6$  in mechanisms that are of functional importance but not essential for binding to a particular muscle-specific DNA sequence.

We used an in vitro selection strategy (9) to test whether such mutations might have more subtle effects on how MyoD binds specifically to DNA. To identify sequences to which

AUTHOR:  
SEE QUERY  
PAGE: \_\_\_\_\_

**BASIC REGIONS:**

MyoD	K R K T T N A D R R K A A T M R E R R R L S K V	++++
E12	Q K A E R E K E R R V A N N A R E R L R V R D I	No
Twist	Q S Y E E L Q T Q R V M A N V R E R Q R T Q S L	ND

**MUTANTS:**

MD (E12B)	[Q K A E R E K E R R V A N N A R E R L R L S K V]	No
MD (E12B-A)	[Q K A E R E K E R R V A A N A R E R L R L S K V]	No
MD (E12B-AT)	[Q K A E R E K E R R V A A T A R E R L R L S K V]	+
MD (E12BJ)	[Q K A E R E K E R R V A N N A R E R L R V R D I]	No
E12 (MDB)	Q K A [T T N A D R R K A A T M R E R R R V R D I]	No
E12 (MDBJ)	Q K A [T T N A D R R K A A T M R E R R R L S K V]	++
E12 (AT, MDJ)	Q K A E R E K E R R V A A T A R E R L R L S K V	+
E12 (AT, K)	Q K A E R E K E R R V A A T A R E R L R V R K I	+
E12 (AT)	Q K A E R E K E R R V A A T A R E R L R V R D I	No

1 5 10 15  
Basic Junction

FIG. 2. Myogenic activity of MyoD and E12 BR and junction mutants. Each of these mutants has been described previously (18, 57), and their sequences are compared with sequences from mouse MyoD, E12, and Twist. Amino acids that are identical to those of MyoD are underlined, positions that are conserved in most bHLH proteins are shaded, and entire BR and junction regions that have been swapped are bracketed. The column at the right indicates the relative activities of these proteins when assayed previously by transfection for conversion of cultured cells into muscle (18, 57); activity is denoted as ++++ (frequency of myogenic conversion obtained with wild-type MyoD), ++ (30 to 50% of that obtained with MyoD), + (5 to 30% of that obtained with wild-type MyoD), No (myogenic conversion was not detected), or ND (not done).

these mutants bind preferentially, we used sequence libraries in which only positions within and flanking the CAN NTG consensus are randomized (Fig. 3A), so that the position of bHLH protein binding along the DNA is fixed. This strategy makes it possible to sequence the selected sites as a pool and thereby to analyze a very large population of selected sites simultaneously (8, 9). It reveals the relative preferences for individual bases at each site position and can detect subtle differences that might not be identified through more conventional approaches.

This assay has previously shown that the preferred MyoD binding consensus is (G/A)ACAGCTG(T/C) (Fig. 3B and C) and that the E2A proteins E12 and E47 overlap considerably with MyoD in their binding properties but prefer sites that have an asymmetric CACCTG core sequence (Fig. 3C) (9). However, in contrast to either of these proteins, the MD(E12B) mutant prefers the sequence (G/A)CCATATGG(T/C), which differs from the MyoD preferred site over the eight central base pairs and contains the distinct core sequence CAT ATG (Fig. 3B and C). This sequence and related elements are normally targeted by the bHLH protein Twist, an E-protein partner that is involved in mesodermal cell fate specification (15, 27, 37, 52, 60) (Fig. 2). Back-substitution of A<sub>5</sub> of MyoD into MD(E12B), which is not sufficient for myogenic activity in cell culture assays (57), results in preferences that are slightly more similar to those of MyoD at positions  $\pm 4$  [MD(E12B-A) (Fig. 2, 3B, and 3C)]. However, introduction of

both A<sub>5</sub> and T<sub>6</sub>, which restores myogenesis (11, 57), results in preferences across the entire site that are indistinguishable from those of MyoD [MD(E12B-AT) (Fig. 2, 3B, and 3C)].

To determine whether these sequence preferences reflect significant differences in binding affinity and specificity, we compared levels of binding of these proteins to individual oligonucleotides that correspond to the MyoD and Twist preferences and differ only at positions within and adjacent to the CAN NTG consensus (Fig. 3D). Supporting the in vitro selection findings, both MyoD and MyoD(E12B-AT) homodimers bound with higher affinity to the preferred MyoD site than to the Twist site (Fig. 3D, lanes 1, 4, 5, and 8). In contrast, the Twist site was preferred by MD(E12B) and, to a lesser extent, MyoD(E12B-A) (Fig. 3D, lanes 2, 3, 6, and 7). In a binding competition assay, specific DNA binding by MD(E12B-AT) was competed much more effectively by the MyoD site (Fig. 4A, lanes 4, 7, 10, 13, and 16), and binding by either MD(E12B) or MD(E12B-A) was competed better by the Twist site (Fig. 4B, lanes 2, 3, 8, 9, 14, and 15). A c-Myc preferred site (CACGTG [not shown]) was a relatively poor competitor of binding by each of these proteins (Fig. 4A and B, lanes 17 to 19). The data show that introduction of A<sub>5</sub> and T<sub>6</sub> into MD(E12B) restores not only myogenic activity (Fig. 2) but also the MyoD DNA binding preference. This substitution affects sequence recognition across 4 bp within each half-site (Fig. 3A and B), indicating a global effect on how the BR helix is positioned on the DNA. The finding that MD(E12B) is distinct

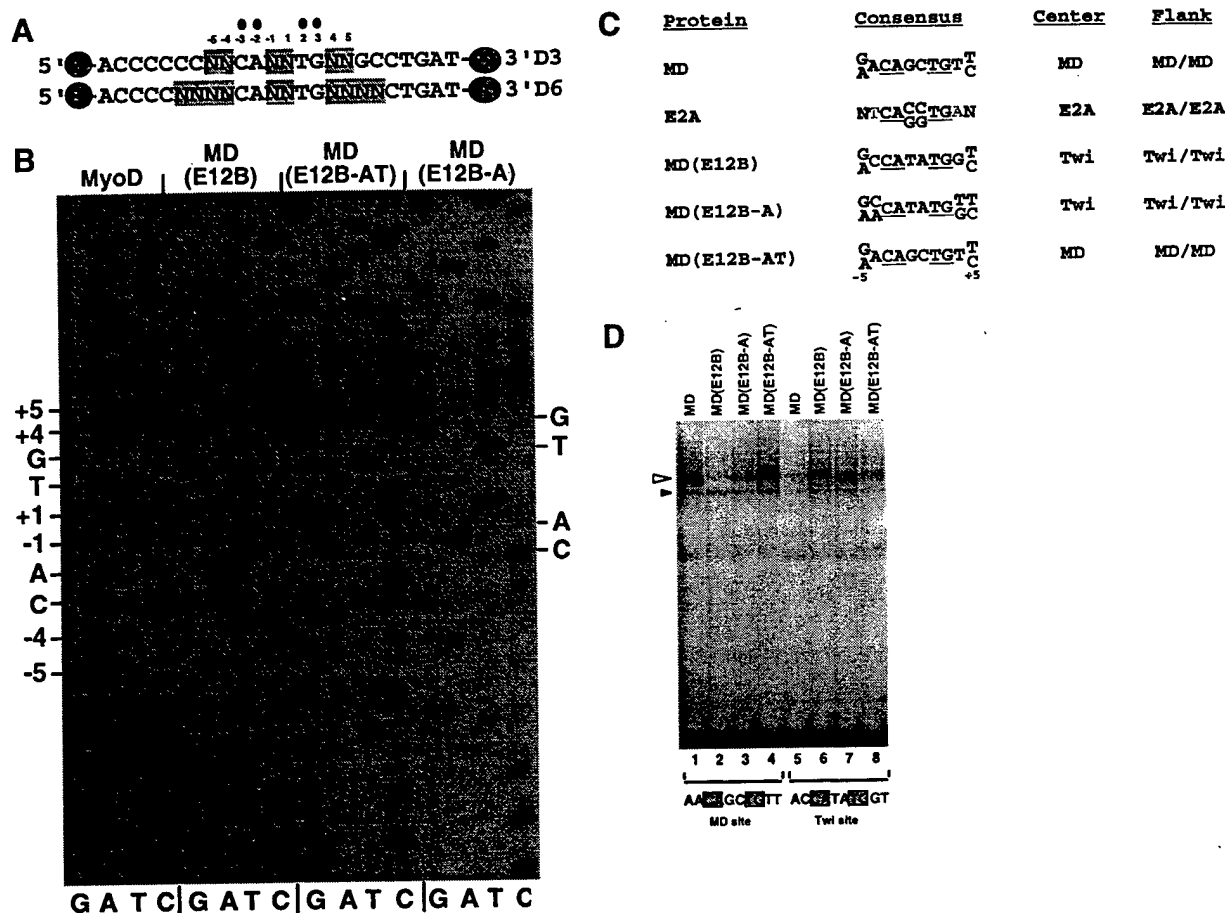


FIG. 3. In vitro selection assay of binding site preferences. (A) Core sequences of the random sequence oligonucleotide libraries D3 and D6 (8, 9). In each library, the bases shown are flanked by sequences which correspond to primers (A and B) that allow selected sequences to be recovered by PCR. A' indicates that primer A corresponds to the opposite strand. (B) Sequences of preferred binding sites. Starting with the D6 oligonucleotide random sequence library (A), three rounds of sequential selection and PCR amplification were performed for binding to the proteins indicated. A sample of the final selected population of binding sites was then sequenced directly as a pool and analyzed by autoradiography. The MyoD preferences at positions  $\pm 1$  described previously (9) are more prominent after additional selection rounds (not shown). (C) Summary of sequence preferences identified by in vitro selection in panel B. MyoD and E2A homodimer preferences were described in reference 9. Binding site positions are numbered as in panel B, and grey letters indicate bases that were selected against. The CAN NTG consensus that was fixed in these experiments is underlined. Twi, Twist. (D) Binding of MyoD BR mutants to individual oligonucleotide sites, which differed only at the sequences shown. In this EMSA, which was analyzed by phosphorimaging, each sample contained the indicated in vitro-translated protein at a concentration of 40 pM and DNA that was labeled to the same specific activity at 550 pM. Specific and background species are indicated by open and closed triangles, respectively.

from either MyoD or E12 in its binding sequence preference also indicates that DNA recognition by an E2A BR can be profoundly influenced by its molecular context.

**Influence of BR positioning on MyoD-E2A and Twist-E2A heterodimer sequence preferences.** Twist and E2A proteins appear to cooperate in vivo to regulate transcription through CAT ATG sites (27), suggesting that the DNA sequence recognition properties of E2A might be altered by heterodimerization with Twist. However, an alternative possibility is that functional Twist-E2A recognition sites are distinct from their in vitro binding preference (28). To address this question, we performed in vitro selection on Twist-E12 complexes. Twist homodimers and Twist-E12 heterodimers both preferred sites that contain the core sequence CAT ATG (Fig. 5A and B). They were similar to MD(E12B) and especially to MD (E12B-A) in their preferences at  $\pm 4$  but selected MyoD-like sequences at  $\pm 5$  (Fig. 3B and 3C, 5A, and 5B). The symmetry of this preferred sequence suggests that in the Twist-E12 protein-DNA complex, the Twist and E12 BRs each prefer the same half-site sequence. In contrast, and as observed previously (9), MyoD-E12 heterodimers selected a MyoD-like half

site at positions +4 and +5, an E2A-like half-site at -4 and -5, and CC or GG bases in the center of the site (Fig. 5A and B), indicating asymmetric binding. Apparently, an E2A BR normally prefers distinct half-sites in the context of these two bHLH dimerization partners, indicating an intermolecular effect on how it interacts specifically with DNA.

To investigate how heterodimer formation influences the binding preferences of the E12 and MyoD BRs, we performed in vitro selection on combinations of MyoD and E12 BR mutants. When the BR of one partner within a MyoD-E12 heterodimer was substituted with that of the other, the heterodimer binding preferences outside the CAN NTG consensus corresponded to those of the individual BRs. For example, unlike MD(E12B) homodimers (Fig. 3B and C), heterodimers of MD(E12B)+E12 preferred wild-type heterodimer sequences in the center of the site, and selected E2A-like sequences in both flanking regions, at  $\pm 4$  and  $\pm 5$  (Fig. 5A and B). A heterodimer of MyoD and an E12 protein containing the MyoD BR [E12(MDB) (Fig. 2A)] similarly selected a wild-type heterodimer preference within the CAN NTG motif but preferred a MyoD-like sequence at  $\pm 4$  and  $\pm 5$

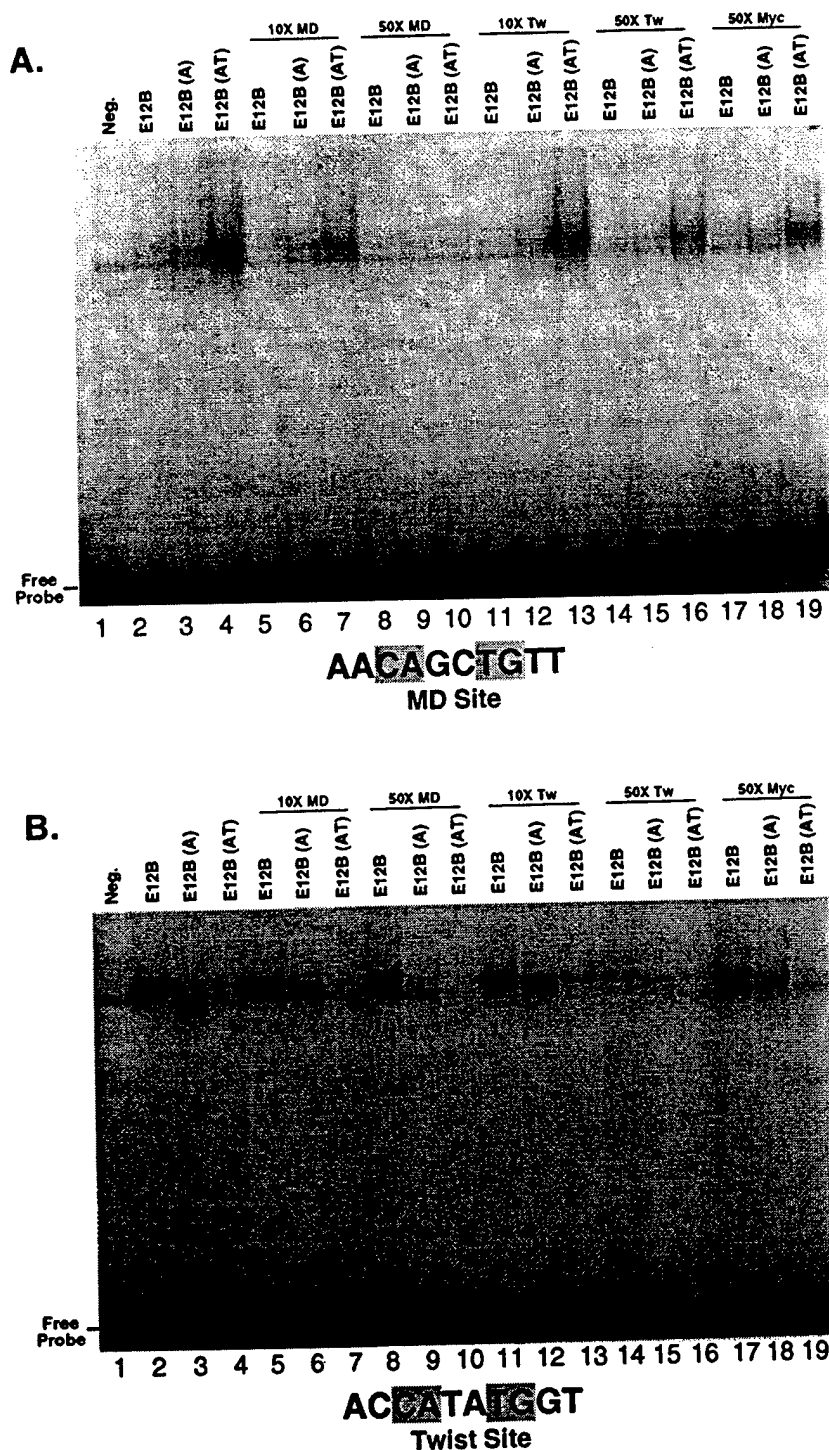


FIG. 4. Specificity of MyoD BR mutant DNA binding. (A) Competition analysis of binding to the labeled MyoD preferred site, analyzed by EMSA and autoradiography. The indicated in vitro-translated proteins and DNA labeled to the same specific activity were present at concentrations of 50 and 900 pM, respectively. When the samples were mixed, unlabeled competitor DNA sites were added at the indicated ratios relative to the labeled probe. Tw, Twist. (B) Competition analysis of binding to the Twist preferred site, performed as for panel A.

(Fig. 5A and B). In contrast, MD(E12B)-E12(MDB) heterodimers had a binding preference more similar to that of Twist (Fig. 5A and B), indicating that placement of each BR in the protein context of the other partner affected binding over the entire site. A striking aspect of our findings is that each of

the mutant homo- or heterodimer protein complexes that we have examined selected sequences that correspond to particular patterns preferred by MyoD, E2A, or Twist protein (Fig. 3C and 5B).

These in vitro selection findings were supported by assays of

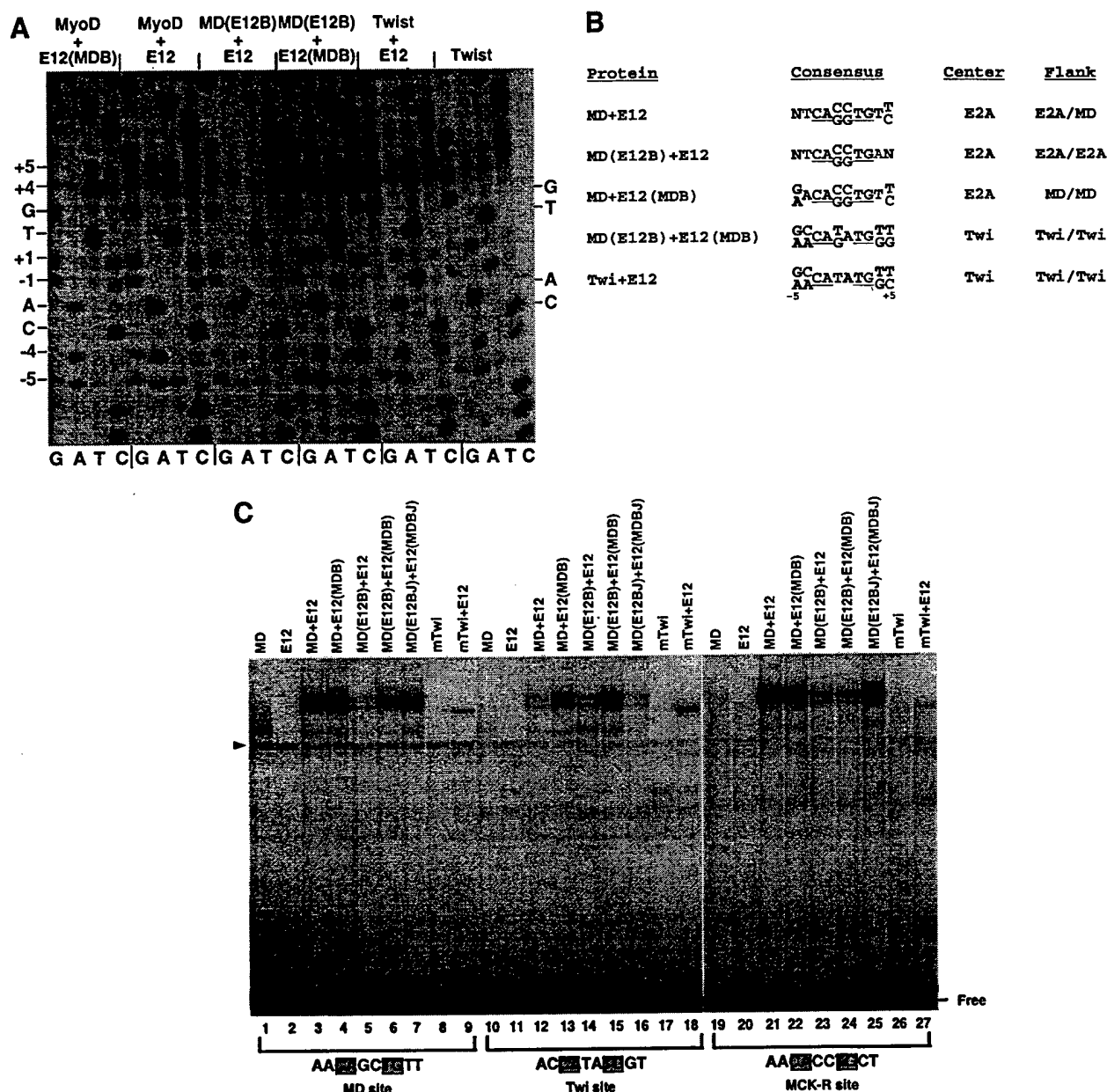


FIG. 5. Binding site preferences of MyoD, E2A, and Twist heterodimer complexes. (A) In vitro selection analysis of binding site preferences. Four rounds of selection from the D3 library (Fig. 3A) were performed as for each in vitro-translated protein complex. In each case, the heterodimer complex could be easily identified in the EMSA on the basis of mobility (9), particularly because E12 homodimers bind DNA poorly (Fig. 9). In the Twist homodimer selection, binding to Twist-E12 heterodimers was selected for in the first round, because of the relatively low level of Twist homodimer binding. Subsequent rounds were performed with Twist homodimers. Each sample was analyzed by sequencing and autoradiography as for Fig. 3B. (B) Summary of sequence preferences identified in panel A, depicted as in Fig. 3C. MyoD-E2A heterodimer preferences were also described previously (9). Twist, Twist. (C) Binding of bHLH heterodimers to individual preferred sites, analyzed by EMSA and phosphorimaging. E2A-derived proteins were present at a concentration of 8 pM, and Twist and MyoD-derived proteins were present at 19 pM. The indicated DNA sites that had been labeled to the same specific activity were present at 550 pM. The MCK-R site differs from the others only at the positions shown. A background species is indicated by a triangle.

binding to individual sites, including a sequence from a muscle-specific regulatory region (MCK-R). This site, which corresponds to the MyoD-E12 heterodimer in vitro binding preference and responds to MyoD in vivo, was used in the original analysis of the myogenic residues (9, 17, 57). In an EMSA, MyoD-E12 heterodimers bound with higher affinity to either the MCK-R or MyoD site than to the Twist site (Fig. 5C, lanes 3, 12, and 21). MyoD(E12B)-E12 heterodimers only slightly preferred the MCK-R heterodimer site to the Twist site but

appeared to prefer either of these sequences to the MyoD site (Fig. 5C, lanes 5, 14, and 23). As the preferences of MD(E12B-A) and MD(E12B-AT) homodimers would predict, introduction of both A<sub>5</sub> and T<sub>6</sub> into MD(E12B) altered its sequence preferences as a heterodimer with E12, so that they were more similar to those of MyoD (not shown). MyoD-E12(MDB) heterodimers only modestly preferred the MyoD or MCK-R site in comparison to the Twist site (Fig. 5C, lanes 4, 13, and 22). In contrast, the Twist site was preferred by



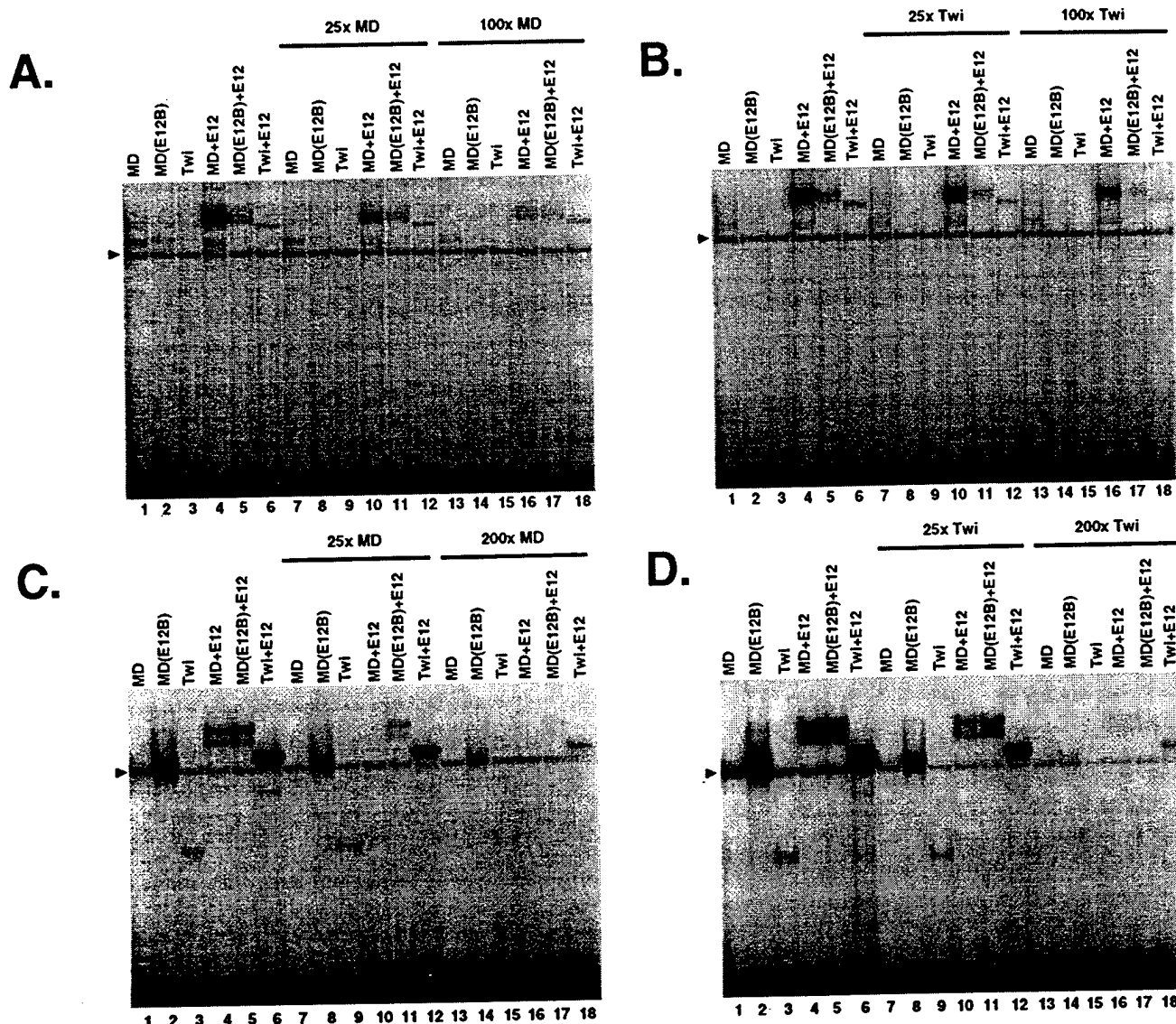


FIG. 6. Binding competition analysis of DNA binding by bHLH heterodimers. (A and B) Binding of the indicated protein complexes to the labeled MyoD site (Fig. 3D) was competed by addition of an unlabeled binding site at ratios indicated above the gel. These experiments were performed and analyzed as for Fig. 4 except that labeled DNA was present at 600 pM, E12 protein present at 8 pM, and all other proteins were present at 19 pM. Twi, Twist. (C and D) Binding of the indicated protein complexes to the labeled Twist site (Fig. 3D) was competed by addition of the indicated unlabeled sites. These experiments were performed described for panel A and B except that labeled DNA was present at 1.1 nM, and they were analyzed by autoradiography. Note that the gel shown in panel C was exposed longer than that shown in panel D, as indicated by comparison of lanes 1 to 6. A background species is indicated by a triangle.

MD(E12B)-E12(MDB), Twist, and Twist-E12 complexes (Fig. 5C, lanes 6, 8, 9, 15, 17, 18, 24, 26, and 27).

Binding site competition and protein titration assays also supported the *in vitro* selection data. The MyoD site competed more effectively than the Twist site for binding by either MyoD homodimers or MyoD-E12 heterodimers (Fig. 6A and B, lanes 1, 4, 7, 10, 13, and 16). In contrast, the Twist site competed more effectively for binding by MD(E12B), MD(E12B)-E12, Twist, and Twist-E12 complexes, although these latter complexes appeared to bind with less specificity than did MyoD-E12 complexes (Fig. 6C and D, lanes 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, and 18). However, the distinct binding specificities of MyoD-E12 and Twist-E12 heterodimers were apparent in a protein titration assay in which the amount of MyoD or Twist protein was varied under conditions of low DNA concentration

(Fig. 7A and B, lanes 1 to 6 and 13 to 18) that more closely represent differences in binding affinity (13). Also in agreement with results described above (Fig. 5C, lanes 14 and 23), heterodimers of MD(E12B) plus E12 bind to the MCK-R site with decreased specificity and with slightly lower affinity than MyoD-E12 complexes (Fig. 7A and B, lanes 7 to 12).

To investigate the role of the BR-HLH junction region in BR positioning, we examined the DNA binding preferences of the MD(E12BJ) and E12(MDBJ) mutants, each of which contains both the BR and junction of the other partner (Fig. 2). In contrast to MD(E12B)-E12(MDB) heterodimers (Fig. 5A and B; Fig. 5C, lanes 6, 15, and 24), MD(E12BJ)-E12(MDBJ) heterodimers (Fig. 2A) bound to the MyoD, Twist, and MCK-R sites with relative preferences that are comparable to those of MD-E12 heterodimers (Fig. 5C, lanes 3, 7, 12, 16, 21,



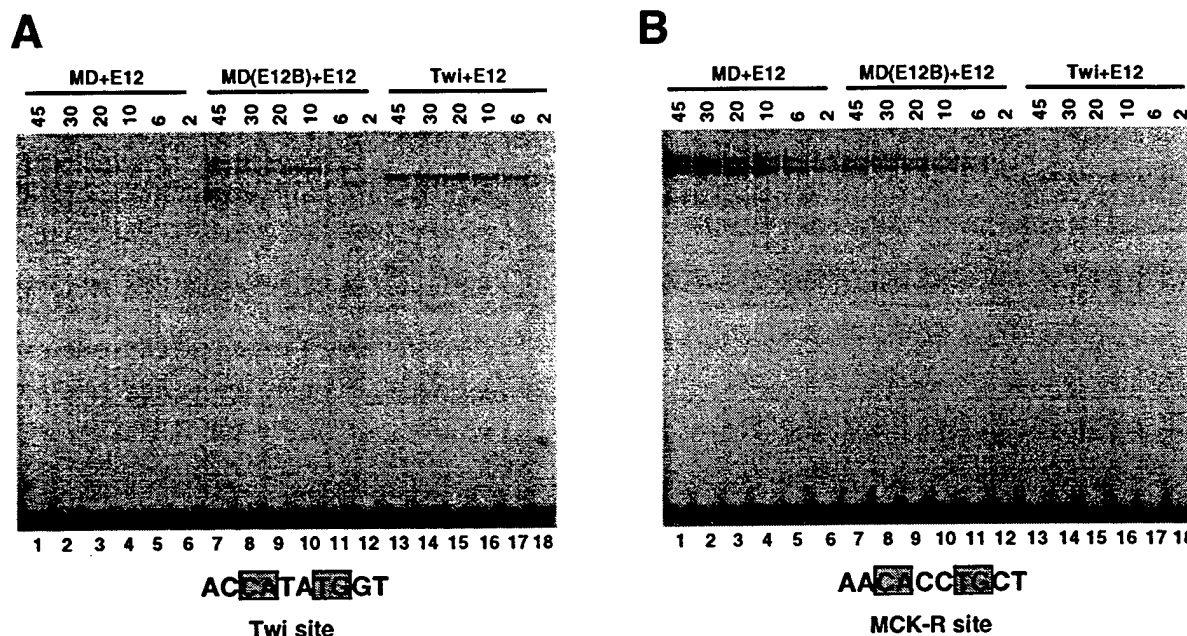


FIG. 7. Protein titration of DNA binding by bHLH heterodimers. (A) Binding to the Twist (Tw) site, analyzed by EMSA and phosphorimaging. In each experiment, E12 was present at 8 pM and DNA that had been labeled to the same specific activity was present at 5 pM. The indicated partner proteins were present at the concentrations (picomolar) shown above the gel. (B) Binding to the MCK-R site, analyzed as for panel A.

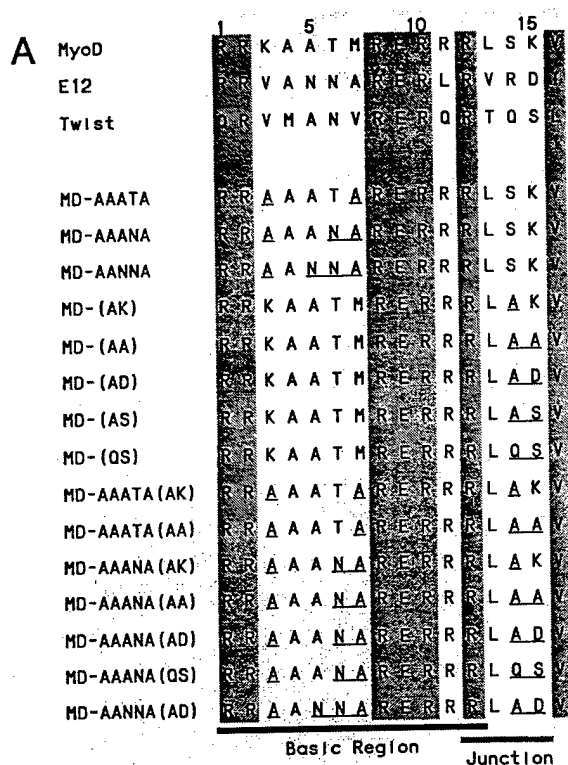
and 25). Apparently, the Twist-like sequence preference resulting from simultaneous mispairing of both the MyoD and E12 BRs (Fig. 5A and B) can be corrected by matching each of these BRs with the corresponding junction region. Similarly, and in contrast to MD(E12B) homodimers, MD(E12BJ) homodimers bind to the MyoD, Twist, and MCK-R sites with preferences that are similar to those of E2A proteins (Fig. 8B and C, lane 20, and data not shown). These findings indicate that the BR-HLH junction can be critical for establishing the sequence specificity of an E2A BR, presumably because it influences how the BR is positioned on the DNA.

**Contributions of the BR and junction to binding affinity and specificity.** It has been shown previously that introduction of  $A_5$ ,  $T_6$ , and either the junction region or  $K_{15}$  of MyoD confers upon E12 the capacity to induce myogenesis (Fig. 2) (18). In the MyoD-DNA complex,  $A_5$  and  $T_6$  are not positioned to allow direct protein-protein contact (Fig. 1) (35), but we have shown that they are critical for the DNA sequence preferences of MyoD, apparently because they affect the conformation of the BR-DNA complex. We have also determined that the junction region can influence how the E2A BR binds DNA. These observations suggest the possibility that the capacity for myogenesis derives entirely from the conformation of the DNA-bound MyoD BR, a model which would predict that the sequence preference of each of these bHLH proteins is established by amino acids at BR positions 5, 6, and 15. We have investigated this model by determining how individual substitutions at these positions, which have been shown to be critical *in vivo*, influence the DNA binding preferences of MyoD.

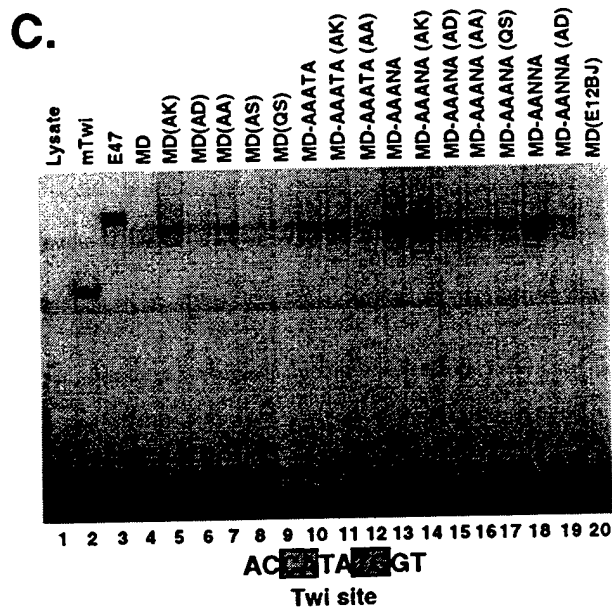
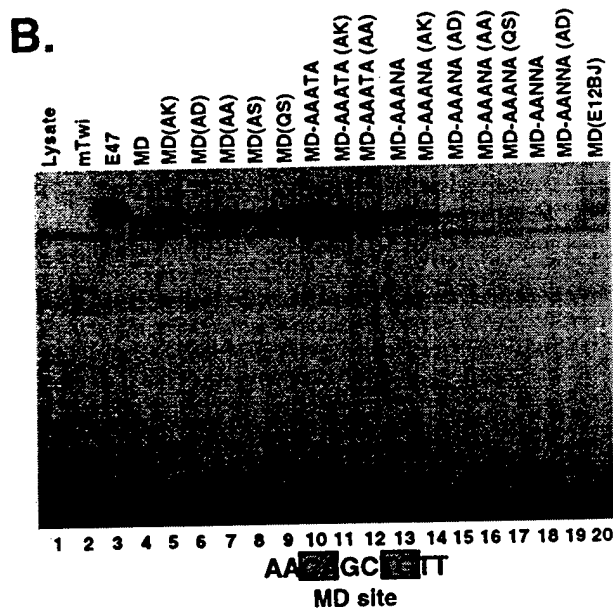
To address the importance of the MyoD junction region for DNA binding, we altered MyoD positions 14 and 15 (Fig. 8A) and left position 13 intact because it is not required for the MyoD sequence preference in the MD(E12B-AT) mutant (Fig. 2 and 3C). Substitution of alanine for  $S_{14}$ , which does not interact with DNA (35), increased binding affinity [MD(AK) (Fig. 8A; Fig. 8B and C, lanes 4 and 5), perhaps by stabilizing

the BR helix. The preference of MD(AK) for the MyoD site was not substantially altered by replacement of position 15 with alanine [MD(AA)] or with either glutamic acid [MD(AD)] or serine [MD(AS) and MD(QS)], which correspond to residues from E12 or Twist, respectively (Fig. 8; Fig. 8B and C, lanes 5 to 9). The relative preferences of these mutants for the MyoD site are comparable to the binding preferences of other proteins that were confirmed by binding competition analysis (Fig. 4 and 6). Apparently, appropriately specific DNA binding by MyoD homodimers is not impaired by a variety of BR-HLH junction substitutions, including nonconservative mutations of  $K_{15}$ . This flexibility contrasts with the importance of the junction region for positioning the E12 BR and with the requirement for  $K_{15}$  for myogenesis.

To investigate the role of BR positions 5 and 6 in a neutral context, we first substituted alanine for two nonconserved BR residues (MD-AAATA [Fig. 8A]) that are not predicted to be required for DNA binding (22, 35). This substitution proportionally increased binding to both sites in the context of MyoD (MD-AAATA [Fig. 8B and C, lanes 10]) and enhanced specificity for the MyoD site in the context of MD(AA) (Fig. 8A; Fig. 8B and C, lanes 12). Replacement of  $T_6$  with asparagine conferred a preference for the Twist site (MD-AAANA [Fig. 8A; Fig. 8B and C, lanes 10 and 13]), a finding that parallels the preferences of MD(E12B-AT) and MD(E12B-A) (Fig. 3B and C). This effect was not diminished by various BR-HLH junction mutations or enhanced by presence of Twist junction residues (Fig. 8B and C, lanes 13 to 17), indicating that  $N_6$  is the most important of these residues for the Twist sequence preference. To test whether E2A amino acids that correspond to the three myogenic residues could specify an E2A-like DNA binding preference, we introduced an asparagine at BR position 7 into MD-AAANA and MD-AAANA(AD), the latter of which contains the  $D_{15}$  residue characteristic of E2A proteins (Fig. 8A). In contrast to MD(E12BJ), these mutants strongly preferred the Twist site to the MyoD or MCK-R sites (Fig. 8B



**FIG. 8.** Effects of bHLH BR and BR-HLH junction residues on MyoD binding preferences. (A) Mutagenesis analysis of the MyoD BR and junction. MyoD BR mutant sequences are compared with the MyoD, E12, and Twist BR sequences (Fig. 2). Conserved bHLH residues are shaded, and residues that are altered within full-length MyoD are underlined. (B) Binding of MyoD mutants described in panel A to the MyoD preferred site. These mutants are compared with the indicated wild-type proteins, and binding is assayed as for Fig. 3D except that each protein is present at 40 pM and DNA labeled to the same specific activity is present at 400 pM. E47 is an alternatively spliced E2A protein that binds DNA well as a homodimer (40). Twi, Twist. (C) Binding of MyoD mutants to the Twist preferred site, assayed as for panel B.



and C, lanes 18 to 20, and data not shown), indicating that establishment of an E2A homodimer sequence preference requires additional E2A BR or junction residues and that the conformational mechanisms that dictate this asymmetric sequence preference might be complex.

In the examples that we have analyzed, MyoD mutants that lack myogenic activity bind preferentially to the Twist site (Fig. 2 and 3C), raising the question of whether changes in DNA binding preferences accompany conversion of E12 to a myo-

genic protein through introduction of MyoD BR and junction residues. E12 homodimers do not bind DNA as well as the E2A protein E47 (Fig. 9, lanes 1, 2, 8, 9, 15, and 16), which also cannot induce myogenesis (18). Introduction of the MyoD BR into E12 is not sufficient for myogenesis [E12(MDB) (Fig. 2)] but sharply increased binding of E12 to all three sites and was associated with a modest preference for the MyoD site (Fig. 9, lanes 3, 10, and 17). The E12(MDBJ) mutant, which can induce myogenesis (Fig. 2), bound to each of the three sites at a

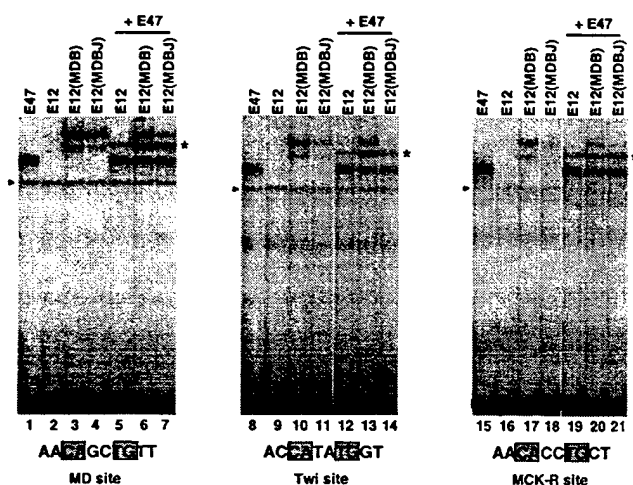


FIG. 9. DNA binding by E12 mutants. DNA binding by the indicated protein complexes is assayed as for Fig. 5C except that all E12 derivatives are present at 8 pM and E47 is present at 19 pM. A protein-DNA complex of intermediate mobility that corresponds to E47-E12 heterodimers is indicated by an asterisk, and a background species is indicated by a closed triangle.

lower level than E12(MDB) and did not have a markedly increased preference for either the MyoD or MCK-R sites (Fig. 9, lanes 4, 11, and 18). Heterodimerization with E47 increased the relative levels with which E12(MDBJ) bound to the MyoD and MCK-R sites (Fig. 9, lanes 6, 7, 13, 14, 20, and 21) but also did not identify DNA binding effects that appear to be sufficient to account for the different functional properties of E12(MDB) and E12(MDBJ). These findings further support the idea that the MyoD junction region is not critical for DNA binding (Fig. 8B and C, lanes 4 to 9) and instead is important for myogenesis because it is involved in other interactions (18).

## DISCUSSION

**bHLH protein DNA binding specificity deriving from effects on BR-DNA conformation.** The myogenic MyoD BR residues  $A_5$  and  $T_6$  are essential for myogenesis but not for binding of MyoD-E2A heterodimers to a muscle-specific site in vitro or in vivo (18, 57). However, we have determined that these residues are required for MyoD to bind DNA with its characteristic specificity for particular CAN NTG sites. Substitution of asparagine for  $T_6$ , and especially for both  $A_5$  and  $T_6$ , results in MyoD binding preferentially to a Twist site (Fig. 8B and C, lanes 10, 13, and 18). The Twist-like MD(E12B) sequence preference is affected partially by substitution of  $A_5$  for the corresponding asparagine [MD(E12B-A) (Fig. 3C)] but is reconfigured by introduction of both  $A_5$  and  $T_6$  so that it is indistinguishable from that of wild-type MyoD [MD(E12B-AT), Fig. 3C]. The data indicate that MyoD residues  $A_5$  and  $T_6$  are each critical for its DNA binding sequence preferences and that the  $N_6$  residue, which is common to the Twist and MD(E12B-A) BRs (Fig. 2), is important for the Twist-like preference. Mutations of these individual BR residues alter sequence preferences across each half-site (Fig. 3C), raising the question of how they might have such a global effect on how the BR helices and the DNA interact preferentially with each other.

A structure of MyoD obtained by X-ray crystallography suggests how  $A_5$  and  $T_6$  might influence binding sequence specificity. When bound to its preferred recognition site, MyoD

does not directly contact base pairs that it specifies in the center of and flanking the CAN NTG consensus (35). However,  $A_5$  and  $T_6$  allow the MyoD BR helix to pack more tightly into the major groove than do the corresponding  $N_5$  and  $N_6$  residues of E2A proteins, in part because of their smaller sizes (Fig. 1 and 2) (35). As a result, the MyoD BR residues  $T_6$  and  $R_2$  directly contact CAN NTG bases at  $\pm 2$  and  $\pm 3$  respectively, and  $R_1$  binds a backbone phosphate at  $\pm 6$  (Fig. 1) (35). In contrast, in E47  $R_2$  swings out of the major groove and contacts the backbone, and the residue at position 1 does not interact directly with the DNA (12, 19). Supporting the idea that  $A_5$  and  $T_6$  influence the conformation of the DNA-bound BR, substitution of asparagine for  $A_5$  in MyoD increases its sensitivity to protease digestion (29). Our findings suggest that protein-DNA interactions that depend specifically on the MyoD  $A_5$  and  $T_6$  residues may directly influence how the BR helix interacts preferentially with the DNA and thereby indirectly specify its characteristic sequence preferences at positions within and flanking the CAN NTG consensus.

Such indirect conformational effects also appear to be critical for the E2A and Twist sequence preferences. When E47 homodimers bind DNA, a single subunit contacts a base in the center of the site through  $R_{10}$  (Fig. 2). This interaction could be important for the asymmetric E2A homodimer sequence preference (19). However, the Twist-like sequence preference that is characteristic of Twist-E2A heterodimers and MD(E12B) homodimers is different across each 5-bp half-site and symmetric (Fig. 3C and 5B), suggesting that it is likely to be established indirectly, through an intermolecular effect that involves a distinct positioning of the BR helix. Introduction of the E12 BR-HLH junction region into MD(E12B) corrects its binding preference so it is like that of E2A homodimers [MD(E12BJ) (Fig. 5C, lanes 7, 16, and 25; Fig. 8B and C, lanes 20)], implicating the BR-HLH junction in this effect. Presumably, the E2A junction acts in concert with the asparagines at BR positions 5 and 6 (Fig. 2), although the Twist-like preference of the MD-AANNA(AD) mutant (Fig. 8B and C, lane 19, and data not shown) suggests that the E2A junction residue  $D_{15}$  is not sufficient. The finding that E2A proteins can be targeted to different DNA sequences by different dimer partners may have important implications for their in vivo functions.

In contrast, the BR-HLH junction region does not have a strong influence on the MyoD DNA binding preference. Various MyoD junction mutations do not substantially diminish its preference for a MyoD site (Fig. 8B and C, lanes 5 to 9). In addition, the similar sequence preferences of E12(MDB) and E12(MDBJ) homodimers (Fig. 9, lanes 3, 4, 10, 11, 17, and 18) contrast sharply with the different specificities of MD(E12B) and MD(E12BJ) (Fig. 3D, lanes 2 and 6; Fig. 8B and C, lanes 20). This apparent difference between MyoD and E2A proteins might derive from the distinct arrangement of the BR helix on the DNA that results from presence of MyoD residues  $A_5$  and  $T_6$ .

It is striking that as a group, these various bHLH mutants and dimer combinations bind DNA with a limited number of discrete sequence preferences (Figs. 3C and 5B). Presumably, each of these preferences reflects a preferred conformational state that is dictated by how each BR helix and the corresponding DNA sequence conform to each other in an induced fit (49). This mechanism for recognizing particular CAN NTG sites appears to be different from the direct recognition of central bases that is characteristic of bHLH proteins that contain  $R_{13}$  and bind to CACGTG or CATGTG sites (20, 21, 48). Consistent with this idea, BR residues 5 and 6 do not appear to be important for the function of the  $R_{13}$ -containing bHLH

protein c-Myc (10). In E2A and its tissue-specific dimerization partners, a more flexible conformation-based mechanism might have evolved to increase adaptability in both sequence recognition and function, so that different combinations of these proteins can result in distinct protein-DNA conformations that correspond to particular DNA sequence preferences. Such a model may be particularly plausible for bHLH proteins, because folding of the BR into an  $\alpha$  helix is driven by its interaction with the DNA (2).

**BR-DNA conformation, DNA binding specificity, and myogenesis.** The observation that the MyoD junction and  $K_{15}$  are not required for an appropriate DNA binding specificity (Fig. 8B and C, lanes 6 to 9; Fig. 9) supports the model that  $K_{15}$  is involved in other essential interactions (18). However, our experiments also pose the question of how the functional importance of  $A_5$  and  $T_6$  might be related to their effects on DNA recognition. Of the MyoD BR mutants that we have analyzed, those that do not induce myogenesis bind to DNA as homodimers with a Twist-like preference [MD(E12B) and MD(E12B-A) (Fig. 2 and 3C)]. Heterodimers of MD(E12B) with E12 prefer a heterodimer site (Fig. 5B), but with markedly diminished specificity compared to MyoD-E12 dimers (Fig. 5C, lanes 3, 5, 12, 14, 21, and 23; Fig. 6; Fig. 7A and B, lanes 1 to 12). This finding suggests that at least in part,  $A_5$  and  $T_6$  may be significant for myogenesis because they restrict the DNA binding specificity of MyoD and other myogenic bHLH proteins, so that they are less likely to bind inappropriate sites. However, other observations support a role for the  $A_5$  and  $T_6$  residues in protein-protein interactions. They have been implicated in binding to other proteins off the DNA (26, 38), and evidence indicates that they are required for activation domain exposure (5, 29, 57) and cooperative DNA binding (3). Finally, unlike MyoD, MD(E12B) can activate transcription of a reporter only in particular cell lines, implicating the BR in protein-protein interactions (57).

In light of evidence that  $A_5$  and  $T_6$  establish the conformation of the DNA-bound BR, it is an attractive model that this effect might influence the function of myogenic bHLH proteins directly, by affecting their interactions with other proteins. Given that relatively subtle alterations of the MyoD BR and junction region can enhance MyoD DNA binding significantly [MD(AK) and MD(AAATA) (Fig. 8B and C, lanes 4, 5, and 10)], it appears likely that cooperative protein-protein interactions with the BR and junction could influence binding affinity. It has been demonstrated recently that MyoD binds cooperatively with other DNA binding proteins to a particular muscle-specific promoter (4). The E box sequences through which MyoD activates transcription in the context of this promoter can differ from those that it binds preferentially *in vitro* (28), suggesting that DNA sequence recognition may be influenced by interactions with cooperating proteins *in vivo*. In addition, interactions with cooperating proteins might be influenced in turn by the specificity of DNA sequence recognition, as suggested by evidence that for MyoD and E proteins, the choice between homo- or heterodimer formation may be dictated by the DNA binding affinities of the individual BRs (36, 59). Our findings are consistent with the idea that deceptively subtle aspects of sequence recognition could be important for the biological activity of MyoD, if they influence functionally critical interactions that might also involve  $K_{15}$  or other MyoD regions.

#### ACKNOWLEDGMENTS

For critically reading the manuscript we thank Robert Davis, Stephen Tapscott, members of the Blackwell laboratory, and Tom Ellenberger, whom we also thank for help with computer graphics. We thank Lau-

ren Snider for advice on bacterial expression and partial purification of MyoD, and we thank Amy Chen and Jonathan Mitchell for contributing to early stages of this project. The *Xenopus* Twist cDNA was provided by Ralph Rupp, and the murine Twist cDNA was donated by Doug Spicer and Andrew Lassar. This work was initiated in the laboratory of Harold Weintraub, to whom T.K.B. is particularly grateful for support, advice, and inspiration.

This work was supported by grant DAMD17-94-J-4063 from the U.S. Army Breast Cancer Program to T.K.B.

#### REFERENCES

- Aksan, I., and C. R. Goding. 1998. Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-box elements *in vitro* and *in vivo*. *Mol. Cell. Biol.* 18:6930-6938.
- Anthony-Cahill, S. J., P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. Stafford, C. Altenbach, W. L. Hubbell, and W. F. DeGrado. 1992. Molecular characterization of helix-loop-helix peptides. *Science* 255:979-983.
- Bengal, E., O. Flores, P. N. Rangarajan, A. Chen, H. Weintraub, and I. M. Verma. 1994. Positive control mutations in the MyoD basic region fail to show cooperative DNA binding and transcriptional activation *in vitro*. *Proc. Natl. Acad. Sci. USA* 91:6221-6225.
- Biesiada, E., Y. Hamamori, L. Kedes, and V. Sartorelli. 1999. Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter. *Mol. Cell. Biol.* 19:2577-2584.
- Black, B. L., J. D. Molkentin, and E. N. Olson. 1998. Multiple roles for the MyoD basic region in transmission of transcriptional activation signals and interaction with MEF2. *Mol. Cell. Biol.* 18:69-77.
- Blackwell, T. K. 1995. Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol.* 254:604-618.
- Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman, and H. Weintraub. 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* 13:5216-5224.
- Blackwell, T. K., L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub. 1990. Sequence-specific DNA binding by the c-Myc protein. *Science* 250:1149-1151.
- Blackwell, T. K., and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* 250:1104-1110.
- Bodis, S., T. Hemesath, and D. E. Fisher. 1997. Highly conserved asparagine in the basic domain of Myc is dispensable for DNA binding, transformation, and apoptosis. *Biochem. Mol. Med.* 60:102-107.
- Brennan, T. J., T. Chakraborty, and E. N. Olson. 1991. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl. Acad. Sci. USA* 88:5675-5679.
- Brownlie, P., T. Ceska, M. Lamers, C. Romier, G. Stier, H. Teo, and D. Suck. 1997. The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* 5:509-520.
- Carey, J. 1991. Gel retardation. *Methods Enzymol.* 208:103-117.
- Chakraborty, T., T. J. Brennan, L. Li, D. Edmondson, and E. N. Olson. 1991. Inefficient homooligomerization contributes to the dependence of myogenin on E2A products for efficient DNA binding. *Mol. Cell. Biol.* 11:3633-3641.
- Cripps, R. M., B. L. Black, B. Zhao, C. L. Lien, R. A. Schulz, and E. N. Olson. 1998. The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during *Drosophila* myogenesis. *Genes Dev.* 12:422-434.
- Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato. 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. USA* 89:599-602.
- Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* 60:733-746.
- Davis, R. L., and H. Weintraub. 1992. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* 256:1027-1030.
- Ellenberger, T., D. Fass, M. Arnaud, and S. C. Harrison. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8:970-980.
- Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363:38-45.
- Ferre-D'Amare, A. R., P. Pognonec, R. G. Roeder, and S. K. Burley. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* 13:180-189.
- Fisher, D. E., L. A. Parent, and P. A. Sharp. 1993. High affinity DNA-binding Myc analogs: recognition by an  $\alpha$  helix. *Cell* 72:467-476.
- Fisher, F., D. H. Crouch, P.-S. Jayaraman, W. Clark, D. A. F. Gillespie, and C. R. Goding. 1993. Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs *in vivo*. *EMBO J.* 12:5075-5082.

24. Gould, K. A., and E. H. Bresnick. 1998. Sequence determinants of DNA binding by the hematopoietic helix-loop-helix transcription factor TAL1: importance of sequences flanking the E-box core. *Gene Expr.* 7:87-101.
25. Halazonetis, T. D., and A. N. Kandil. 1992. Predicted structural similarities of the DNA binding domains of c-Myc and endonuclease Eco RI. *Science* 255:464-466.
26. Hamamori, Y., H. Y. Wu, V. Sartorelli, and L. Kedes. 1997. The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. *Mol. Cell. Biol.* 17:6563-6573.
27. Harfe, B. D., C. S. Branda, M. Krause, M. J. Stern, and A. Fire. 1998. MyoD and the specification of muscle and non-muscle fates during postembryonic development of the *C. elegans* mesoderm. *Development* 125:2479-2488.
28. Huang, J., T. K. Blackwell, L. Kedes, and H. Weintraub. 1996. Differences between MyoD DNA binding and activation site requirements revealed by a functional random sequence selection. *Mol. Cell. Biol.* 16:3893-3900.
29. Huang, J., H. Weintraub, and L. Kedes. 1998. Intramolecular regulation of MyoD activation domain conformation and function. *Mol. Cell. Biol.* 18:5478-5484.
30. Jennings, B. H., D. M. Tyler, and S. J. Bray. 1999. Target specificities of *Drosophila* enhancer of split basic helix-loop-helix proteins. *Mol. Cell. Biol.* 19:4600-4610.
31. Kophengnavong, T., A. S. Carroll, and T. K. Blackwell. 1999. The SKN-1 amino-terminal arm is a DNA specificity segment. *Mol. Cell. Biol.* 19:3039-3050.
32. Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* 66:305-315.
33. Lee, J. E. 1997. Basic helix-loop-helix genes in neural development. *Curr. Opin. Neurobiol.* 7:13-20.
34. Lemercier, C., R. Q. To, R. A. Carrasco, and S. F. Konieczny. 1998. The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of myoD. *EMBO J.* 17:1412-1422.
35. Ma, P. C., M. A. Rould, H. Weintraub, and C. O. Pabo. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77:451-459.
36. Maleki, S. J., C. A. Royer, and B. K. Hurlburt. 1997. MyoD-E12 heterodimers and MyoD-MyoD homodimers are equally stable. *Biochemistry* 36:6762-6767.
37. Michelson, A. 1996. A new turn (or two) for Twist. *Science* 272:1449-1450.
38. Molkentin, J. D., B. L. Black, J. F. Martin, and E. N. Olson. 1995. Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell* 83:1125-1136.
39. Molkentin, J. D., and E. N. Olson. 1996. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc. Natl. Acad. Sci. USA* 93:9366-9373.
40. Murre, C., P. S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56:777-783.
41. Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 58:537-544.
42. Neuhold, L., and B. Wold. 1993. HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* 74:1033-1042.
43. Olson, E. N., and W. H. Klein. 1994. bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev.* 8:1-8.
44. Pabo, C. O., and R. T. Sauer. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053-1095.
45. Parraga, A., L. Belloso, A. R. Ferre-D'Amare, and S. K. Burley. 1998. Cocystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* 6:661-672.
46. Postigo, A. A., and D. C. Dean. 1999. ZEB represses transcription through interaction with the corepressor CtBP. *Proc. Natl. Acad. Sci. USA* 96:6683-6688.
47. Schwarz, J. J., T. Chakraborty, J. Martin, J. M. Zhou, and E. N. Olson. 1992. The basic region of myogenin cooperates with two transcription activation domains to induce muscle-specific transcription. *Mol. Cell. Biol.* 12:266-275.
48. Shimizu, T. X., A. Toumoto, K. Ihara, M. Shimizu, Y. Kyogoku, N. Ogawa, Y. Oshima, and T. Hakoshima. 1997. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J.* 16:4689-4697.
49. Spolar, R. S., and M. T. Record, Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777-784.
50. Steitz, T. A. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q. Rev. Biophys.* 23:205-280.
51. Studier, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219:37-44.
52. Szymanski, P., and M. Levine. 1995. Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *EMBO J.* 14:2229-2238.
53. Thayer, M. J., and H. Weintraub. 1993. A cellular factor stimulates the DNA-binding activity of MyoD and E47. *Proc. Natl. Acad. Sci. USA* 90:6483-6487.
54. Van Antwerp, M. E., D. G. Chen, C. Chang, and E. V. Prochownik. 1992. A point mutation in the MyoD basic domain imparts c-Myc-like properties. *Proc. Natl. Acad. Sci. USA* 89:9010-9014.
55. Voronova, A., and D. Baltimore. 1990. Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. *Proc. Natl. Acad. Sci. USA* 87:4722-4726.
56. Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar. 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* 251:761-766.
57. Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott. 1991. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* 5:1377-1386.
58. Weintraub, H., T. Genetta, and T. Kadesch. 1994. Tissue-specific gene activation by MyoD: determination of specificity by cis-acting repression elements. *Genes Dev.* 8:2203-2211.
59. Wendt, H., R. M. Thomas, and T. Ellenberger. 1998. DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J. Biol. Chem.* 273:5735-5743.
60. Yin, Z., X. L. Xu, and M. Frasch. 1997. Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* 124:4971-4982.